THE ROLE OF POLYHOMEOTIC IN POLYCOMB GROUP-MEDIATED REGULATION OF THE TUMOUR SUPPRESSOR P16^{INK4A}

Lucas Antoine de Breed

Thesis submitted towards the degree of

Doctor of Philosophy

University College London

30 September 2009

Molecular Oncology Laboratory
CRUK London Research Institute
Lincoln's Inn Fields, London

Faculty of Life Sciences
University College London
Gower Street, London

Declaration

I, Lucas Antoine de Breed, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Lucas Antoine de Breed Date

Abstract

The *INK4A* tumour suppressor gene is negatively regulated by the Polycomb Group (PcG) complexes, which plays an important role in stem cell- and cancer biology. Polycomb Repressive Complex 1 (PRC1) members posterior sex comb protein BMI1 and polycomb proteins CBX7 and CBX8 have been shown to be directly involved in repression of *INK4A* and delay of senescence, whereas evidence from mouse genetics suggests that polyhomeotic protein Ph2 is also required. However, it is still unknown why the mammalian PRC1 complex has so many possible permutations compared to the basic system in *Drosphila* and how this relates to the regulation of a single target gene. The aim was therefore to define the requirement and function of the three human polyhomeotic members, HPH1, -2 and -3, for INK4A repression. It was found that depletion of all HPH proteins by shRNA-mediated knockdown in primary human fibroblasts results in derepression of *INK4A* and premature senescence. In contrast, overexpression of HPH proteins leads to activation of INK4A accompanied by loss of PRC1 binding and PcG-associated histone H3K27 trimethylation at the locus. These diametrically opposite findings suggest that a fine balance in HPH protein levels is required for PcG repression. Furthermore, biochemical investigation of the interaction between CBX and HPH proteins did not demonstrate any level of specificity, but emphasised the RNA binding capacity of both proteins, which was further investigated. In addition, by tandem affinity purification and mass spectrometry analysis it was found that HPH2 interacts with both PRC1 components and novel proteins, which were studied for involvement in *INK4A* repression. In summary, this thesis demonstrates an essential role for the HPH proteins in the regulation of *INK4A* and provides novel insights into their functional contribution to PcG repression.

Acknowledgements

First and foremost, I would like to thank my supervisor, Gordon Peters, for giving me the opportunity to join his group and for mentoring me while pursuing my scientific ambitions. I have highly appreciated your near unrestricted availability and have greatly enjoyed our stimulating discussions. I will look back on my time in your lab as a great learning curve, from which I have gained both scientific rigour and the ability to experimentally put my ideas into practice.

I am also grateful to the other members of the Molecular Oncology lab, with whom it was both educational and fun working together. I would like to particularly acknowledge Marc Rodriguez Niedenführ's contribution to the data discussed in this thesis. Thank you for your limitless help in getting the tools to move the project forward. I hope you have also become fascinated by the polyhomeotics and will be carrying the project forward from hereon. Furthermore, I am thankful for the help of Hollie Chandler, Emma Anderton, Julie Stock and Helen Pemberton in reviewing this document.

For research support, I am indebted to the Graham Clark and his team at the LRI Equipment Park, who have processed countless samples on my behalf and have always been very helpful. Also I am grateful to the support of LRI Cell Services, without whom I would not have been able to perform my protein purification experiments. Similarly, Mark Skehel and his colleagues of the Protein Identification and Proteomics group at Clare Hall deserve noteworthy acknowledgement for mass spectrometry analysis. I also would like to thank Nicola O'Reilly and her group for immunogenic peptide synthesis.

I also would like to thank my thesis committee, Julie Cooper and Julian Downward, for their ideas and advice, particularly in the early stages of the project. Furthermore, I would like to thank Sally Leevers and the LRI Research Administration team for providing guidance and support over the course of my PhD.

Lastly, I would not have been able to complete this work without the emotional support of my family, friends and Sarah, who closely shared both the frustrations and joys over the course of this work, but always provided support and strong encouragement, for which I am very grateful.

Table of Contents

Declaration....2
Abstract....3
Acknowledgements....4
Table of Contents....5
Figures and Tables....9
Abbreviations....11
1 Introduction....15
1.1 The Polycomb Group Family....15
1.1.1 *Drosophila* Polycomb Biology....15
1.1.2 Polycomb Complexes....16
1.1.3 Mechanism of PcG-mediated Repression....20
1.2 PcG in Stem Cells and Cancer....28
1.2.1 PcG proteins in stem cell self-renewal....28
1.2.2 PcG proteins in cancer....30
1.3 PcG proteins in regulation of the INK4A tumour suppressor....32
1.3.1 The *INK4B-ARF-INK4A* locus and its products....32
1.3.2 Role of the *INK4A-ARF-INK4B* locus in senescence....34
1.3.3 Different Types of Senescence....35
1.3.4 *INK4B-ARF-INK4A,* senescence and tumour suppression....36
1.3.5 Senescence and ageing....37
1.3.6 PcG-mediated regulation of *INK4A*....39
1.4 Characterising PRC1 complexes that regulate INK4a....41
1.4.1 HPH Family overview....41
1.5 Thesis Aims....46
2 Materials and Methods....47
2.1 Solutions....47
2.2 DNA techniques....53
2.2.1 Plasmid vectors....54
2.2.2 PCR....54
2.2.3 Restriction digestion, ligation and transformation....55
2.2.4 Plasmid DNA preparation....56
2.2.5 DNA sequencing....57

2.3 Cell culture 58
2.3.1 Cell strains/lines 58
2.3.2 Storage, recovery and harvesting of cells 59
2.3.3 Transfection of human cell lines 60
2.3.4 Viral transduction of human cells 60
2.4 Protein biochemistry 61
2.4.1 Preparation of total cell lysate 61
2.4.2 SDS-PAGE 61
2.4.3 Immunoblotting 62
2.4.4 Immunoprecipitation 64
2.4.5 Chromatin Immunoprecipitation 64
2.4.6 Tandem Affinity Purification 66
2.4.7 Protein gel staining 67
2.4.8 LC/MS/MS Mass spectrometry 67
2.4.9 Recombinant protein expression 68
2.4.10 In vitro binding assay 70
2.5 RNA techniques 70
2.5.1 RNA isolation 70
2.5.2 Reverse transcription quantitative PCR (RT-qPCR) 70
2.5.3 Absolute quantification using qPCR 73
2.5.4 Sense/Antisense-specific RT-qPCR 73
2.5.5 RNA ChIP 74
2.5.6 RNA interference 74
2.6 Cell Biology Techniques 76
2.6.1 Senescence-associated β-Galactosidase Assay 76
3 The Generation of Antibodies against HPH2 and CBX7 77
3.1 Introduction 77
3.2 Generation of CBX7 antibodies 77
3.2.1 Mouse monoclonal antibodies to a CBX7 fragment 77
3.2.2 Human recombinant antibodies to CBX7 fragments 78
3.2.3 Rabbit polyclonal antibodies to CBX7 peptides 79
3.3 Generation of HPH2 Antibodies 80
3.3.1 Human recombinant antibodies to full length HPH2b 80
3.3.2 Rabbit polyclonal antibodies to HPH2 peptides 81
3.4 Polyclonal antibodies against other HPH proteins 82

3.5 Discussion 82
4 The Characterisation of the Polyhomeotic Family and the Interaction with the Human PC Family 100
4.1 Introduction 100
4.2 Characterization of the HPH Family 100
4.3 Characterization of the interaction of the HPH family with the CBX family 102
4.3.1 All HPH family members bind CBX7 102
4.4 All human PC proteins bind to HPH2 103
4.5 Discussion 106
5 The Role of the Polyhomeotic Family in the Regulation of *INK4A* 120
5.1 Introduction 120
5.2 HPH2 knockdown in HDFs results in derepression of INK4a 121
5.3 HPH2 binds to the promoter and exon 1 region of INK4A 122
5.4 HPH2 depletion does not affect repression of INK4A by Cbx7 123
5.5 Ink4a repression by Cbx7 is incapacitated in Mph2$^{-/-}$ MEFs 124
5.6 Overexpression of HPH2 in HDFs 125
5.6.1 Involvement of HPH1 and HPH3 in INK4A repression 126
5.7 Discussion 127
6 The Functional Role of the Human Polyhomeotic Proteins in the Regulation of *INK4A* 141
6.1 Introduction 141
6.2 Identification of novel Interactions of HPH2 141
HPH2 interactions with RNA 146
6.2.1 Profiling of antisense transcription along the *INK4B-ARF-INK4A* locus 146
6.2.2 HPH2 binds transcripts antisense to the *INK4A* promoter 147
6.3 Discussion 148
7 Discussion 163
7.1 Lessons about PRC1 composition 163
7.2 Interactions between PcG complexes 165
7.3 The HPH proteins and cell proliferation 166
7.4 Novel interacting partners for HPH2 168
7.4.1 Splicing and RNA binding 169
7.4.2 Higher-order chromatin regulation 171
7.4.3 DNA damage response 172
7.5 Conclusions 172

References.....174

Figures and Tables

Figure 1.1 Classical model of PcG-mediated repression.....23
Figure 1.2 Model for PcG involvement in normal differentiation and opportunities for tumourigenesis 31
Figure 1.3 The human *INK4B-ARF-INK4A* locus and its products 33
Figure 1.4 Conservation of polyhomeotic family in different species.....42
Figure 3.1 Expression of CBX7 Fragments.....86
Figure 3.2 Overview of the CBX7 antigens 88
Figure 3.3 CBX7 Recombinant Human Antibodies Evaluation.....90
Figure 3.4. CBX7 Polyclonal Antibodies Evaluation 92
Figure 3.5 Overview of HPH2 Antigens. 94
Figure 3.6 HPH2 Recombinant Human Antibodies Evaluation.....96
Figure 3.7 HPH2 Polyclonal Antibodies Evaluation. 98
Figure 4.1 Potential specificity of the interaction between CBX7 and HPH2b (adapted from (Nicholls, 2006)).....108
Figure 4.2 The characterisation of HPH transcripts in human cells 110
Figure 4.3. CBX7 interacts with both isoforms of HPH2. 112
Figure 4.4. CBX7 interacts with all three HPH proteins.....113
Figure 4.5 Preferential binding of HPH2 to endogenous CBX7. 114
Figure 4.6 CBX7 interacts with the FCS domain of HPH2 115
Figure 4.7 HPH2 interacts with all five human PC proteins. 116
Figure 4.8 Chromodomain mutations do not affect the interaction of CBX7 with HPH2 117
Figure 4.9 HPH proteins interact with a highly conserved domain of CBX7.....118
Figure 5.1 HPH2 is involved in the regulation of human *INK4/ARF* 129
Figure 5.2 HPH2 co-localises with other PRC1 components at the *INK4B-ARF-INK4A* locus. 131
Figure 5.3 HPH2 knockdown does not affect *INK4A* repression by mouse Cbx7 133
Figure 5.4 Cbx7 repression of Ink4a is impaired in Mph2-/- MEFs.....135
Figure 5.5 HPH2 overexpression activates *INK4A* by depletion of PcG-mediated repression 137
Figure 5.6 All HPHs are involved in *INK4A* regulation.....139
Figure 5.7. Depletion of two HPH proteins suggests synergy in repression.....140

Figure 6.1 Putative hypothesis for the regulation of splicing in the context of PcG complexes 149
Figure 6.2 Evaluation of HPH2 Tandem Affinity Purification 151
Figure 6.3. Identifcation of HPH2a interacting proteins 152
Figure 6.5 Sense/antisense-specific RT-qPCR Methodology 159
Figure 6.6 Profile of *INK4B-ARF-INK4A* sense and antisense transcription 160
Figure 6.7. HPH2 co-immunoprecipitates RNA transcripts antisense to the *INK4A* promoter region 161
Figure 6.8 Mutation of the FCS RNA binding domain of HPH2 does not affect its induction of *INK4A* 162

Table 1. PcG proteins representing the core components of PRC1 and PRC2 in *Drosophila*, mouse and human (the names commonly used in this thesis are in bold). 19
Table 2. Commonly used sequencing primers 58
Table 3. List of commonly used primary antibodies detailing their species, type and origin/source 64
Table 4. List of primers used for amplification of cDNA from mature mRNA 72
Table 5. List of primers used for amplification of the *INK4/ARF* locus. 72
Table 6. List of primers used for differential PCR of HPH transcripts 73
Table 7 siRNA sequences that were cloned into a shRNA format in either a retroviral or lentiviral vector or both 76
Table 8 List of peptides used for generation of rabbit polyclonal antibodies 85
Table 9 Reproducible identifications of proteins co-purifying with TAP-HPH2a 153
Table 10 Single identifications of proteins co-purifying with TAP-HPH2a 158

Abbreviations

°C.	degrees centigrade
Ad5	Adenovirus 5
ARF	alternative reading frame
ATP	adenosine 5'-triphosphate
BMI1	B-lymphoma Moloney MLV insertion region-1
bp	base pair
BSA	bovine serum albumin
C-terminus	carboxy terminus
CBP	calmodulin-binding peptide
CBX	chromobox homolog
CDK	cyclin-dependent kinase
cDNA	copy DNA
CMV	cytomegalovirus
CRUK	Cancer Research UK
DDR	DNA damage response
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxyribonucleoside 5'-triphosphate
DSB	double strand break
DTT	dithiothreitol
E. coli	*Escherichia coli*
ECL	enhanced chemiluminescence
EDR	early developmental-regulated
EDTA	Ethylenediaminetetraacetic acid
ERK	extracellular signal-regulated kinase
FBS	foetal bovine serum
FCS	Phenylalanine Cysteine Serine
FDF	fetal dermal fibroblast
g	gravity (centrifugal force)
G1/G2	Gap phase 1/2 (of the cell division cycle)

Gag	group antigen gene
GFP	green fluorescent protein
GS4B	glutathione sepharose 4B
GST	glutathione-S-transferase
HA tag	hemagglutinin epitope tag
HBS	HEPES-buffered saline
HD1	homology domain 1
HDAC	histone deacetylase
HDF	human diploid fibroblast
HIS	histidine
HOX	homeobox
HP	heterochromatin protein
HPH	human polyhomeotic
HRP	horseradish peroxidase
HS68	human skin 68
hTERT	human telomerase reverse transcriptase
Ig	Immunoglobulin
INK4	inhibitor of CDK4
IP	immunoprecipitation
IPTG	isopropyl-beta-D-thiogalactopyranoside
IVT	in vitro transcription/translation
kb	kilobase
kDa	kilo-Dalton
LRI	London Research Institute of Cancer Research UK, formerly known as Imperial Cancer Research Fund (ICRF)
M	molar
M-phase	mitosis-phase (of the cell division cycle)
MAPK	mitogen-activated protein kinase
MAPKAP	MAPK activated protein
MBLR	Mel18- and Bmi1-like RING finger protein
MCS	muliple cloning site
Mdm2	murine double minute 2
MEF	mouse embryonic fibroblast
MEK	MAPK/ERK kinase
MEL18	melanoma nuclear protein 18

MK	MAPKAP kinase
MLV	murine leukemia virus
Mph	mouse polyhomeotic
N-terminus	amino-terminus
ncRNA	non-coding RNA
NP40	nonidet P40 (ocetylphenoxy polyetoxy ethanol)
PAGE	polyacrylamide gel electrophoresis
PBS-A	phosphate buffered saline A
Pc	polycomb
PcG	polycomb group
PCR	polymerase
PD	population doubling
PHC	polyhomeotic homolog
Pho	pleiohomeotic
PMSF	phenylmethylsulphonyl fluoride
PRC1	polycomb repressive complex 1
PRC2	polycomb repressive complex 2
PSC	posterior sex combs
Rae28	Retinoic acid-activated early-28
RAS	rat sarcoma
RB	retinoblastoma
RING	really interesting new gene
RNA	ribonucleic acid
RNAi	RNA interference
Rpm	revolutions per minute
RT	reverse transcription
RT-qPCR	reverse transcription quantitative PCR
S phase	(DNA) synthesis phase (of the cell division cycle)
SAM	sterile alpha motif
SCM	sex combs on the midleg
SDS	sodium dodecyl sulphate
SET	Suppressor of variegation-Enhancer of zeste-Trithorax
SF3b	splicing factor 3b
shRNA	short hairpin RNA
TAP	tandem affinity purification

TEMED	N,N,N’,N’-tetramethylethylenediamine
TRIS	Tris(hydroxymethyl) aminomethane
TrxG	Trithorax group
Tween-20	Polyoxyethylenesorbitan monolaurate
UK	United Kingdom
v/v	volume to volume
VSV-G	vesicular stomatitis virus G
w/v	weight to volume
XIST	X-inactive-specific transcript
rt	room temperature, 25°C
sec	second
min	minute
h	hour
O/N	overnight
g	gram
m	milli
µ	micro
n	nano
p	pico
l	litre

1 Introduction

1.1 The Polycomb Group Family

The Polycomb Group (PcG) represents a family of transcriptional repressors that play an important role in maintaining the repressed state of genes in the developmental regulatory transcription programs. PcG proteins are considered to be part of the cellular memory system, which ensures the transmission of cellular identities through cell division. Moreover, deregulated expression of PcG proteins has important implications in cancer and stem cell self-renewal (discussed in more detail in section 1.2). Most of the fundamental knowledge about PcG repression was derived from research in *Drosophila.*

1.1.1 *Drosophila* Polycomb Biology

PcG genes were initially discovered in the fruitfly *Drosophila melanogaster*. Mutations in the PcG genes were found to cause multiple homeotic transformations, which were similar to gain-of-function mutations of the homeotic selector genes. The first mutant had a phenotype of disrupted segmentation displaying multiple sex combs, and hence was named *Polycomb* (Sandler et al., 1968). The *Polycomb* gene was subsequently discovered to be part of a gene family, the polycomb group (PcG), controlling segmentation by repression of the homeotic genes (Lewis, 1978). The homeobox or *Hox* genes are a set of transcription factors that are expressed in a spatially restricted manner along the anteroposterior axis during development, which results in the morphological differences between segments from head to tail (Krumlauf, 1994). Consequent identification of the trithorax group (TrxG) that activate homeotic genes has led to a concept of a balance between silencing and activating factors that together provide cellular memory at different stages of development (Kennison and Tamkun, 1988; Shao et al., 1999). In early stage embryonic development *Hox* genes are initially regulated by transiently expressed segmentation genes, but later PcG and TrxG are required to maintain the correct pattern of *Hox* gene expression.

Subsequently identified genetic interactions and biochemical studies indicated that both the members of the PcG and TrxG groups form large multiprotein complexes (reviewed in (Schuettengruber et al., 2007)). These complexes interact directly with the chromatin at their target gene. In *Drosophila*, specific DNA regulatory elements, named Polycomb and Trithorax response elements (PRE and TRE respectively) were found to be necessary and sufficient for recruitment of both PcG and TrxG complexes. Although these elements serve a common function throughout the *Drosophila* genome, the PREs and TREs do not carry any consensus sequence (reviewed in (Schwartz and Pirrotta, 2007)). Recently, the first example of a vertebrate PRE was discovered in mouse (Sing et al., 2009). Nevertheless, it is important to note that there are substantial differences between the *Drosophila* and the mammalian mechanism of PcG repression. For example, four of the DNA binding elements important in *Drosophila* PcG recruitment, GAF, Pipsqueak, Zeste and DSP1, are not conserved in vertebrates (reviewed in (Sparmann and van Lohuizen, 2006)). It is therefore important to be cautious in generalising mechanistic findings about PcG-mediated repression across species. Perhaps one of the most striking differences between *Drosophila* and mammalian Polycomb biology is the expansion in the number and diversity of proteins that constitute the Polycomb complexes.

1.1.2 Polycomb Complexes

The textbook version of PcG-mediated repression holds that the PcG proteins function through the combined activity of two Polycomb repressive complexes (PRC); the initiator complex, PRC2, and the maintenance complex, PRC1 (reviewed in (Schuettengruber et al., 2007)). Although biochemical analysis of both complexes suggests that they are independent entities with separate functions and timing (Ng et al., 2000; Sewalt et al., 1998; Shao et al., 1999; Tie et al., 2001; van Lohuizen et al., 1998), a transient interaction between components of both complexes has been observed in early embryonic development (Poux et al., 2001; van Lohuizen et al., 1998). Furthermore, it has been reported that complex formation is important for the stabilization of PcG proteins both *in vivo* and *in vitro* (Leeb and Wutz, 2007; Montgomery et al., 2005; Pasini et al., 2004). It was found that if one protein is lost, the others are also affected post-transcriptionally (Montgomery et al., 2005). This would

suggest that PcG proteins not incorporated into complexes may be degraded (reviewed in (Schwartz and Pirrotta, 2008)).

It has also been observed that PcG complexes form subnuclear silencing compartments, or Polycomb/PcG bodies, that may play a role in the stable repression of transcription (Saurin et al., 1998). Immunofluorescence of PcG proteins shows clear speckles, but the number of speckles decrease during development (Grimaud et al., 2006). It was shown that that the formation of these Polycomb bodies in *Drosophila*, requires colocalisation of components of the RNA interference (RNAi) machinery (Grimaud et al., 2006). This is particularly interesting due to the increasing observations of RNA involvement in PcG-mediated repression, which will be discussed in more detail in section 1.1.3.5.

1.1.2.1 The PRC1 Complex

The PRC1 complex was first purified from *Drosophila* where it was found to actively inhibit the activity of TrxG complexes (Shao et al., 1999). PRC1 was found to consist of four conserved core components, Polycomb (Pc), Polyhomeotic (Ph), Posterior sex combs (Psc) and Sex combs extra (Sce). Over recent years, it has been established that the mammalian PcG family has multiple paralogues for each *Drosophila* component, with five polycombs (CBX2/4/6/7/8), three polyhomeotic (HPH1/2/3), six posterior sex combs (PSC1-6) and two sex combs extra (RING1A/1B) proteins (Francis et al., 2001; Levine et al., 2002; Whitcomb et al., 2007). At present, there is no universally accepted nomenclature for these proteins. For completeness, all the names used for *Drosophila*, mouse and human PcG proteins are listed in Table 1, where the names in bold will be those used in this thesis. The functional basis of this evolutionary expansion of PRC1 genes is unknown, but within the context of a core complex of four subunits, this diversity could render a large number of permutations of the PRC1 complex in mammalian cells (reviewed in (Gil and Peters, 2006; Whitcomb et al., 2007)). In addition to the four core subunits of PRC1, a number of other proteins have been reported to associate with the complex. Although the *Drosophila* sex-comb on the midleg (Scm) protein is generally considered to be part of PRC1, it has only been found to associate with PRC1 in sub-stoichiometric amounts (Levine et al., 2002; Saurin et al., 2001; Shao et al., 1999). For this reason, Scm is not considered to be a core component of PRC1. In addition, our laboratory has not been able to co-purify either of the two

human homologues of SCM, SCMH1 and SCMH2, with either Cbx7, MEL18 or BMI1 (Maertens et al., 2009).

1.1.2.2 The PRC2 Complex

In comparison to PRC1, evolution has only provided a modest expansion of the genes that contribute to the formation of PRC2 and so the human complex is largely constituted of direct homologues of the *Drosophila* core components. As summarised in Table 1, these include two homologous of the enhancer of zeste (EZH1/2), the homologue of extra sex combs embryonic ectoderm development (EED) and the homologue of suppressor of zeste (SUZ12) (Cao et al., 2002; Satijn et al., 2001). Genetic experiments in mice have illustrated the importance of an integral PRC2 complex, as ablation of any of the PRC2 genes was found to produce early embryonic lethality (O'Carroll et al., 2001; Pasini et al., 2004; Schumacher et al., 1996; Tie et al., 2007). This can be attributed to the fact that all proteins are required for targeted and efficient methyltransferase activity *in vivo* (see below). Additional components associated with the PRC2 complex are the retinoblastoma binding proteins 4 (RBBP4/RBAP48/NURF55) and -7 (RBBP7/RBAP46). Together with SUZ12, these two proteins were shown to be required for the association of EZH2 with nucleosomes (Kuzmichev et al., 2002; Nekrasov et al., 2005). Additional variation in complex formation is established by the different isoforms of EED or expression level of EZH2, which play a role in directing the enzymatic activity of the complex (Kuzmichev et al., 2004; Kuzmichev et al., 2005). This variability in the PRC2 complex has also been observed in *Drosophila* based on alternative Esc subunits (Ohno et al., 2008).

1.1.2.3 The Pho/YY1 complex

In *Drosophila*, a third PcG complex has been proposed, termed PhoRC. PhoRC includes the PRE-binding proteins Pleiohomeotic (Pho) or Pho-like (Phol) that are involved in the recruitment of the PRC2 and PRC1 complexes to target genes by providing sequence-specific DNA binding ((Wang et al., 2004b), reviewed in (Schuettengruber et al., 2007)). Homologues of Pho and Phol have been identified in human, where they have been called Yin-yang 1 and 2 (YY1/YY2) respectively

(Nguyen et al., 2004; Shi et al., 1991). However, no YY1-based complex has yet been identified in mammals and until very recently the existence of a mammalian PRE was questioned, which together have limited the understanding of the role of these proteins in the mammalian mechanism of PcG-mediated repression.

PRC1 (Maintenance)		
Drosophila	Mouse	Human
Polycomb (Pc)	**Cbx2**/M33/pc **Cbx4**/Pc2/PC2 **Cbx6** **Cbx7** **Cbx8**/Pc3	**CBX2**/M33/PC **CBX4**/PC2/hPC2 **CBX6** **CBX7** **CBX8**/PC3/HPC3
Polyhomeiotic (Ph)	**Mph1**/Phc1/Rae28/Edr1 **Mph2**/Phc2/Edr2 **Mph3**/Phc3/Edr3	**HPH1**/PHC1/EDR1/RAE28 **HPH2**/PHC2/EDR2/PH2 **HPH3**/PHC3/EDR3
Sex combs extra (Sce)	**Ring1a**/Ring1 **Ring1b**/Ring2	**RING1A**/RING1/RNF1 **RING1B**/RING2/RNF2
Posterior sex combs (Psc)	Pcgf1/**Nspc1** Pcgf2/**Mel18**/Rnf110/Zfp144 **Pcgf3**/Rnf3/RNF3A Pcgf4/**Bmi1**/Rnf51 **Pcgf5**/Rnf159 Pcgf6/**Mblr**/Rnf134	PCGF1/**NSPC1**/RNF68 PCGF2/**MEL18**/RNF110 **PCGF3**/RNF3/RNF3A PCGF4/**BMI1**/RNF51 **PCGF5**/RNF159 PCGF6/**MBLR**/RNF134

PRC2 (Initiation)		
Drosophila	Mouse	Human
Extra sex combs (Esc)	**Eed**/I7Rn5/lusk	**EED**/WAIT1
Enhancer of zeste (E(z))	**Ezh1** **Ezh2**/Enx-1/KMT6	**EZH1** **EZH2**/ENX-1/KMT6
Suppressor of zeste (Su(z))	**Suz12**	**SUZ12**/CHET9/JJAZ1

Table 1. PcG proteins representing the core components of PRC1 and PRC2 in *Drosophila*, mouse and human. The names commonly used in this thesis are in bold.

1.1.3 Mechanism of PcG-mediated Repression

1.1.3.1 Introduction to Chromatin and Transcription

In eukaryotes, genomic DNA is packaged into a chromatin fiber, which involves the wrapping of the DNA double strand around an octamer of four pairs of the core histones H2A, H2B, H3 and H4. The chromatin fiber is a highly dynamic structure, which is altered in a number of fundamental DNA-based processes, including replication, DNA repair, cell division and transcription. Around 150bp of DNA is wrapped around a single nucleosome, but the density and locations of nucleosomes varies. This density determines the consequent openness or compaction of regions of chromatin and this has an important regulatory role for transcription. In addition, enzymatic modification of the amino terminal tails of the histone proteins, which include acetylation, methylation, phosphorylation, ubiquitination, sumoylation, citrullination, deimination, ADP-ribosylation and proline isomerisation, play an important part in the regulation of transcription. Of these the small covalent modifications acetylation, methylation and phosphorylation are best characterised. It has become increasingly evident that different combinations of histone modifications influence transcriptional regulation of genes and therefore it has been suggested that they constitute an epigenetic- or histone code, which is superimposed on the genetic code (reviewed in (Goldberg et al., 2007)).

Transcription of RNA consists of distinct phases during which RNA polymerase II (Pol II) is recruited to a gene promoter, initiates transcription, elongates the RNA transcript, and finally terminates transcription. Transcription starts with the formation of the preinitiation complex (PIC) at promoters of activated genes. The PIC is composed of three categories of transcription factors, being transactivators, cofactors and basal factors, which work coordinately to recruit PolII and trigger RNA synthesis. An important level of regulation of transcription is the rate of PIC formation, which is also influenced by enhancers at more distal sites of chromatin. In the transcription initiation process, the cofactor and basal factors are generally viewed as the universal machinery, whereas the transactivators are the positive or negative regulators that can work in a cell- or gene-specific manner. Nevertheless, an increasing number of studies suggest that there are also cell type-specific cofactors and basal factors (reviewed in (D'Alessio et al., 2009)).

Over recent years it has been found that for transcription factors to access and bind their target sequences in promoters, local reorganisation of nucleosomal structures is required, and that this is a dynamic and continuous process (reviewed in (Hager et al., 2009)). This reorganisation is orchestrated by the combined effort of chromatin remodelers and chromatin modifiers, of which the latter can be directly recruited by transcription factors. The traditional view that promoters are either in an open or closed chromatin state has been superseded by observations that most promoters are in fact found to be in a wide spectrum of intermediate chromatin states, between these two contrasting extremes (reviewed in (Cairns, 2009)).

Also during transcription elongation, the PolII complex needs to pass through nucleosome structures and therefore recruits chromatin remodelling enzymes that can temporarily disassemble and reform the chromatin structures (Carrozza et al 2005). In addition, different histone modifications are correlated with either active or repressed transcription and can influence the elongation ability of the PolII complex and thereby stall transcription before this phase (Core and Lis, 2008; Stock et al., 2007). The acetylation of histone H3 on lysine 9 (H3K9ac), the di- or trimethylation of histone H3 on lysine 4 (H3K4me2 or –me3), the dimethylation of histone H3 on lysine 36 (H3K36me2) and the ubiquitination of histone H2B at lysine 120 (H2BK120ub) are generally correlated with active transcription. In contrast, trimethylation of histone H3 on lysine 9 and 27 (H3K9me3 and H3K27me3), trimethylation of histone H4 on lysine 20 (H4K20me3) and ubiquitination of H2A on lysine 119 (H2AK119ub) are all linked to transcriptional repression and the possible formation of heterochromatin. However, it is important to recognise that the meaning of these modifications is highly dependent on the context of other modifications. As such, H3K9 methylation for example can both have an active or repressive effect (Vakoc et al., 2005). PcG-mediated repression involves the recognition and establishment of a number of these key chromatin marks, as will be discussed in more detail below.

1.1.3.2 Classical Model of PcG Repression

In the current models of PcG-mediated repression (Figure 1.1), the PRC2 complex interacts with histone deacetylases (HDAC) via EED to remove the acetylated H3K9 mark from transcriptionally active chromatin (Kuzmichev et al., 2002; Tie et al., 2001; van der Vlag and Otte, 1999). In addition, the histone methyltransferase activity of

PRC2 converts the lysine 27 of histone H3 (H3K27) to a trimethylated form. This constitutes a unique enzymatic property of the PRC2 complex and is therefore widely regarded as the hallmark of PcG-mediated repression (Cao et al., 2002; Czermin et al., 2002; Muller et al., 2002). This mark is then recognized by the chromodomain of the Polycomb protein(s) and facilitates binding of the PRC1 complex (Czermin et al., 2002; Fischle et al., 2003; Min et al., 2003). Following binding, PRC1 catalyses the ubiquitination of histone H2A on lysine 119 (H2AK119ub), an essential modification for transcriptional repression (Wang et al., 2004a). In mice, it was shown that the Ring1B E3 ligase protein is the main catalytic subunit, whereas both Ring1A and Bmi1 have been identified as important enhancers of enzymatic activity (Buchwald et al., 2006; Cao et al., 2005).

This model has however been challenged by the observations of PRC1 binding to nucleosomal arrays without histone tails, by the absence of nucleosomes at PcG binding sites and by the PRC2-independent function of PRC1 in X-chromosome inactivation (Francis et al., 2004; Papp and Muller, 2006; Schoeftner et al., 2006). Furthermore, this model does not explain how PRC2 is recruited in the mammalian system, nor how the binding of PRC1 results in stable repression. Therefore, to get a more precise understanding of PcG-mediated repression, it is important to consider the various elements of PcG repression in more detail and in the context of recent findings.

1.1.3.3 Histone Methylation in PcG-mediated repression

Histone H3K27 trimethylation is the hallmark of PcG-mediated repression and is assumed to function as the landing platform for PRC1 recruitment. Binding of PRC1 to chromatin is indeed lost following depletion of PRC2, and the interaction between Polycomb and H3K27me3 can be competed away from *Drosophila* polytene chromosomes, at least in part, by using a H3K27me3 peptide (Bracken et al., 2007; Ringrose et al., 2004). Moreover, genome-wide binding studies in *Drosophila*, mouse and human have found that H3K27me3 is generally correlated with PcG protein binding (Boyer et al., 2006; Bracken et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006). These studies found that the H3K27me3 mark is highly localised to gene promoters at some loci, but spread over large domains at others, such as observed for the *HOX* gene clusters. However, PRC1 was not found to follow the same distribution pattern (Schwartz et al., 2006).

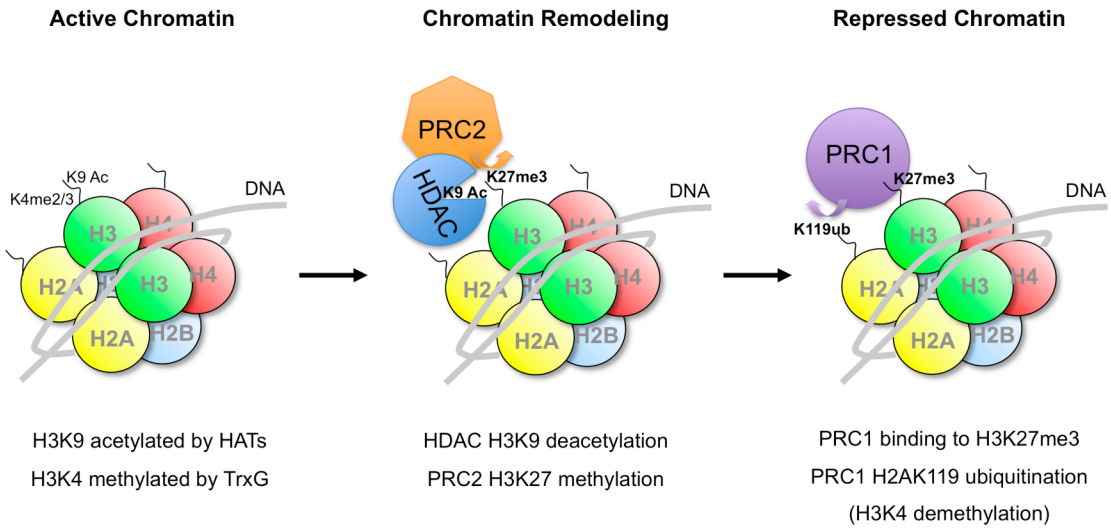


Figure 1.1 Classical model of PcG-mediated repression.

Actively transcribed chromatin typically has acetylated H3K9 (H3K9ac) and di- or trimethylated H3K4 (H3K4me2/me3). Upon initiation of PcG-mediated repression, the PRC2 complex recruits a histone deacetylase (HDAC) to the target chromatin to catalyse the removal of the H3K9ac mark. In addition, PRC2 catalyses the trimethylation of H3K27 (H3K27me3). In the next step, the PRC1 complex recognises and binds the H3K27me3 mark. The ubiquitin E3 ligase activity of the complex then mediates the ubiquitination of H2AK119 (H2AK119ub). It is proposed that H3K4me2/3 is also removed at this stage. These collective events result in the stable repression of transcription.

In fact, in *Drosophila* only a fraction of the H3K27me3 domains are bound by PcG, whereas the overlap in mammals was found to be much higher (reviewed in (Schwartz and Pirrotta, 2008)). Even more so, PRC1 has been reported to be recruited to certain targets in the absence of any H3K27 methylation (Schoeftner et al., 2006; Vincenz and Kerppola, 2008). These observations thus undermine the direct link between methylated H3K27 and PRC1 chromatin binding.

Conversely, the methylation of H3K27 could also be involved in functions independent of PRC1. It has been shown that mono- and dimethylation of H3K27 are also mediated by PRC2 proteins, although the function of these modifications is unknown, as they are not strongly bound by the chromodomain of CBX proteins (Ebert et al., 2004). Ezh2 and Suz12 are only required for the di- and trimethylated forms of

H3K27, whereas Eed is necessary for all stages of H3K27 methylation (Montgomery et al., 2005; Pasini et al., 2004). Interestingly, Ezh1 was found to be responsible for the mono-methylated form of H3K27, and PRC2 complexes incorporating Ezh1 are mainly active in non-proliferating cells and have a robust ability to compact chromatin (Margueron et al., 2008; Shen et al., 2008).

The methylation of H3K27 can also be reversed by the action of the Jumonji C-domain histone demethylases, such as UTX and JMJD3 (reviewed in (Swigut and Wysocka, 2007)). The association of UTX with the Trithorax MLL H3K4 methyltransferases, the localisation of UTX at HOX gene promoters and the direct involvement in HOX gene expression strongly suggest that H3K27 demethylation acts to counteract PcG repression (Cho et al., 2007; Issaeva et al., 2007; Lan et al., 2007; Lee et al., 2007). It is thus believed that the balance between TrxG activation and PcG repression is in part executed by opposing activities of the PRC2 complex and the JmjC-domain demethylases (Swigut and Wysocka, 2007).

PRC2 variants also methylate other target sites than H3K27. It was found that a different isoform of the EED protein of PRC2 directs the methyltransferease activity of the complex to the trimethylation of histone H1 on K26. This mark is recognised by heterochromatin protein 1 (HP1) and could be important for higher order chromatin organisation (Daujat et al., 2005; Kuzmichev et al., 2004). Trimethylation of H3K9 and H4K20 have been found to accompany H3K27me3 on PcG-repressed genes and these marks are involved in the formation of constitutive pericentric heterochromatin (Czermin et al., 2002; Papp and Muller, 2006; Schotta et al., 2004). Although, the trimethylation of H3K9 is normally catalysed by the SUV39H1 methyltransferase, PRC2 has also been found to trimethylate H3K9 *in vitro* (Bannister et al., 2001; Czermin et al., 2002; Kuzmichev et al., 2002; Lachner et al., 2001; Rea et al., 2000). It has also been shown that the chromodomains of a number of murine Pc/Cbx proteins have high affinity for H3K9me3 *in vitro* (Bernstein et al., 2006b). This mark might therefore be involved in a connection between PcG-mediated repression and the induction of heterochromatin formation (Fischle et al., 2003).

1.1.3.4 Ubiquitination in PcG-mediated repression

Although, purified PRC1 complexes have demonstrable ubiquitin ligase activity specific for histone H2A, it is clear that the main components are Psc and Sce or its

homologues that all carry the characteristic RING domain. In the mammalian system, RING1B is widely accepted as the main catalyst of PRC1-mediated ubiquitination of histone H2A (Wang et al., 2004a). However, it was shown that Ring1A can substitute for Ring1B E3 ligase activity (de Napoles et al., 2004; Endoh et al., 2008; Leeb and Wutz, 2007). Further evidence has indicated that there are distinct complexes consisting of only Ring1B and co-factor BMI1 or Mel18, which can directly catalyse the ubiquitination of H2A (Buchwald et al., 2006; Elderkin et al., 2007; Li et al., 2006). A similar complex, also including the KDM2 histone demethylase, was recently found in *Drosophila* and human cells (Gearhart et al., 2006; Lagarou et al., 2008). It is therefore unclear whether ubiquitination of histone H2A is performed by the PRC1 complex or by derivatives of it.

In addition to the ubiquitination of histone H2A, deubiquitination of histone H2B has also been shown to contribute to PcG-mediated repression (van der Knaap et al., 2005). Ubiquitinated histone H2B is required to establish the activating mark H3K4me3 and in *Arabidopsis* it was found that deubiquitination of H2B is required for the methylation of H3K9 and important for the maintenance of heterochromatic silencing (Sridhar et al., 2007). This would thus suggest that the deubiquitination of H2B might be an important link between PcG-mediated repression and more stable silencing by the formation of heterochromatin.

1.1.3.5 Non-coding RNA in PcG-mediated repression

Several observations have implicated non-coding RNA (ncRNA) molecules in PcG-mediated repression, but mechanistic concepts about how ncRNA is employed still need to be elucidated. Firstly, X-inactivation, needed to equal the X-chromosome dosage between sexes, involves both PcG-mediated repression silencing and ncRNA (Silva et al., 2003). Interestingly, X-inactivation requires the ncRNA X inactive-specific transcript (*XIST*) (Brown et al., 1991). Based on the recent identification that *Xist* binds to the PRC2 complex, it was postulated that *XIST* thereby recruits PcG complexes to the X-chromosome destined for inactivation (Zhao et al., 2008). However, it was recently shown that initiation, but not maintenance, of X-inactivation can occur in the absence of *XIST* (Kalantry et al., 2009). This would thereby suggest that PcG-mediated repression would be involved as a secondary mechanism in X-inactivation, possibly facilitating stable silencing of the X chromosome.

In contrast, studies in *Drosophila* report that non-coding transcription through the PRE is required to counteract silencing (Schmitt et al., 2005). A mammalian counterpart was also found by investigation of intergenic ncRNA transcription that accompanies co-linear activation of the human *HOXA* cluster (Sessa et al., 2007). Another report studying all four human *HOX* clusters identified and using the H3K4me3 and H3K36me2 marks as a signature for active transcription identified many conserved regions of ncRNA transcription that co-ordinately expressed with respect to their 3' proximal *HOX* gene (Rinn et al., 2007). This approach also led to the discovery of a transcript regulating the repression of a distal *HOX* gene *in trans* by interacting with the PRC2 complex. Genome-wide follow up studies in different cell types have now identified several thousand of these long ncRNAs of which more than a quarter were found to interact with PRC2 (Guttman et al., 2009; Khalil et al., 2009).

Further work in *Drosophila* has established that components of the RNAi machinery are involved in PcG silencing. It was demonstrated that AGO1, an essential component of the RNAi machinery, recruits EZH2 to mediate the transcriptional silencing of promoters targeted by small interfering RNA (siRNA) (Kim et al., 2006). Furthermore, another study showed that RNAi components are required for long-distance interactions of interchromosomal regions silenced by PcG complexes, but that mutations of key components of the RNAi machinery did not find an essential role in the PcG mechanism (Grimaud et al., 2006). In line with a structural role of RNA in PcG silencing, the ncRNA Kcnq1 overlapping transcript 1 (*Kcnq1ot1*) and PcG proteins are required for the formation of discrete repressive nuclear compartments for imprinting of the *Kcnq1* locus (Pandey et al., 2008; Terranova et al., 2008). Similarly, ncRNA involvement in the imprinting of the Igf2r/h19 domain by long-range interchromosomal looping of PcG proteins has been observed (Li et al., 2008; Sleutels et al., 2002). It thus remains to be established whether ncRNA plays a general role in preventing promoter access to PcG complexes, in directing PcG silencing or in establishing higher order chromatin structures, or all of the above.

1.1.3.6 Repression by stable binding of PRC1

It was long believed that the stable binding of PRC1 to genomic loci blocks the access to transcription factors and the transcriptional machinery and would thereby repress its expression (Cao et al., 2002). However, this has been contrasted by observations in

Drosophila that components of the general transcription factor TFIID can still bind to PcG-repressed genes and interact with PRC1 proteins (Breiling et al., 2001; Saurin et al., 2001). Examination of the regulation of *Ubx*, a *Drospophila* Hox gene, also concluded that PRC1 binding does not prevent transcription factor binding, but prevents the activation of these factors (Dellino et al., 2004). Both the PcG and TrxG complexes were found to bind the the *Ubx* promoter, irrespective of its transcriptional state (Papp and Muller, 2006). This study also observed that *Ubx* activation leads to the loss of H3K27me3 from the promoter and coding region, where it was replaced by H3K4me3, which suggests that only local chromatin marks correlate with transcription. These findings therefore suggest that the mediators of transcription are closely linked with transcriptional repression by PcG.

1.1.3.7 DNA Hypermethylation

In addition to histone modifications, one of the well-known mechanisms of stable gene repression is the methylation of the DNA itself, specifically of CpG islands associated with gene promoters. Methylation-dependent silencing of tumour suppressor genes is commonly observed in and contributes to the development of cancer (reviewed in (Ballestar and Esteller, 2008). Recently, DNA hypermethylation has been directly linked to PcG-mediated repression. DNA methyltransferases (DNMTs) interact directly with EZH2 (Vire et al., 2006) and the presence of both these proteins and EED are required for recruitment of BMI1 to Polycomb bodies (Hernandez-Munoz et al., 2005). In addition, the methylated DNA binding protein (MBD1) interacts with CBX4 and RING1B at heterochromatin foci and these interactions are required for *HOXA* repression (Sakamoto et al., 2007), whereas CBX4 also sumoylates DNMT3a (Li et al., 2007). It has also been found that PcG-mediated methylation of H3K27me3 marks genes for de novo methylation in cancer cells, leading to permanent silencing of target genes (Bracken et al., 2006; Schlesinger et al., 2007).

It thus remains to be determined how these various aspects of the PcG-mediated repression mechanism fit together, what their contribution is and how they are regulated. There is thus a need for further biochemical and functional characterisation of the PRC complexes and their components. Increased understanding of the PcG-repression mechanism will lead to a better appreciation of the roles of these proteins in development, as well as tissue homeostasis and malignancy.

1.2 PcG in Stem Cells and Cancer

Although a number of PcG target genes were already known from classical genetic studies, more recent genome-wide screens of PcG protein binding sites in mouse embryonic stem (ES) cells and human embryonic fibroblasts identified more than a thousand potential target genes (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). Functionally, these genes are related to cell proliferation, cancer, genomic imprinting and X-inactivation, but the most important concept to emerge was that PcG proteins are important for maintaining pluripotency of embryonic stem cells and the identity and self-renewal capacity of somatic stem cells (reviewed in (Ringrose, 2007). On the other hand, disrupted levels of PcG proteins have increasingly become linked to aberrant proliferation in cancer (reviewed in (Gil et al., 2005; Sparmann and van Lohuizen, 2006; Valk-Lingbeek et al., 2004). The morphological defects resulting from PcG mutations can be directly connected to defective stem cell maintenance, whereas expression of PcG proteins is found to be upregulated in cancer cells. The PcG proteins therefore may represent an important element in the 'cancer stem cell hypothesis', which holds that the driving subpopulation of cells in a tumor has stem cell properties (reviewed in (Gupta et al., 2009; Reya et al., 2001)).

1.2.1 PcG proteins in stem cell self-renewal

The first indications that PRC1 proteins play a role in stem cell maintenance came from studies on Bmi1-deficient mice. These animals display haematopoietic- and neurological defects that are caused by a failure in self-renewal of the corresponding stem- and progenitor cell populations (Lessard and Sauvageau, 2003; Leung et al., 2004; Molofsky et al., 2003; Park et al., 2003; van der Lugt et al., 1994). In addition to Bmi1, Mel18, Mph1 and Cbx2/M33 are also highly expressed in haematopoietic progenitor cells and their mutation gives rise to various defects in the haematopoietic system (Akasaka et al., 1997; Core et al., 1997; Ohta et al., 2002; Park et al., 2003; Takihara et al., 1997). The fact that the observed effects on haematopoiesis are different between these mutants, may suggest that the proteins have distinct roles in this process.

Due to the early embryonic lethality of PRC2 mutants, their involvement in adult stem cells has not been well characterised, although Ezh2 was found to control B cell development in conditional knockout mice (Su et al., 2003). However, recent

studies indicate that ES cells lacking *Ezh2* or *Suz12* are viable and proliferate in culture (Chamberlain et al., 2008; Pasini et al., 2007). Even though the lack of core PRC2 components leads to the global loss of H3K27 methylation, H3K27me3 and PcG silencing were maintained at a select number of key developmental target genes by the activity of Ezh1 (Shen et al., 2008). The interaction between Ezh1 and Eed can maintain ES cell self-renewal, but not pluripotency, which additionally equires Suz12 and Ezh2 (Ezhkova et al., 2009; Pasini et al., 2007).

Genome-wide analyses of histone modifications of mouse and human ES cells has found that key regulatory genes of developmental pathways are bound by PcG complexes, have broad H3K27me3 domains and the active H3K4me3 mark at the transcription start sites (Boyer et al., 2006; Lee et al., 2006). This epigenetic state has been termed bivalent and is thought to be characteristic of genes poised for expression (Bernstein et al., 2006a). Upon ES cell differentiation, either the repressive or active mark is lost as a consequence of activation or silencing of the respective gene. Nevertheless, a number of bivalent domains are sustained after differentiation (Barski et al., 2007; Mikkelsen et al., 2007). It was recently discovered that bivalent domains are maintained by a low level of basal transcription, of which elongation is restricted by Ring1b-mediated ubiquitination of H2A (Stock et al., 2007). This thus strongly implicates the PcG proteins in controlling the bivalency of genes.

PcG-mediated repression thus has an important role in ES cells by maintaining self-renewal, pluripotency and priming differentiation, whereas they are required maintaining the identity and self-renewal of somatic stem cells. The significance of PcG proteins in pluripotency is further supported by the fact that SOX2 and OCT3/4, two of the four ‘Yamanaka factors’ used to derive induced pluripotent stem (iPS) cells, and the downstream self-renewal factor NANOG are co-localised with PcG proteins on the promoters of target genes (Boyer et al., 2006; Lee et al., 2006; Takahashi and Yamanaka, 2006), and OCT3/4 was shown to direct the binding of PRC1 in ES cells (Endoh et al., 2008).

PcG proteins in cancer

The first link between PcG and cancer was drawn from the discovery of Bmi1 as a target of insertional mutagenesis in murine leukaemia virus-induced lymphomas (Haupt et al., 1991; van Lohuizen et al., 1991). The name Bmi1 (B-cell Moloney Integration) reflects these observations and it was shown that the gene collaborates with oncogenic Myc (Eμ-Myc) in tumourigenesis. Human BMI1 has since been found to be highly expressed in leukaemia and lymphoma, colorectal cancer, liver cancer, breast cancer, medulloblastoma and glioma, and non-small cell lung cancer (Bea et al., 2001; Bruggeman et al., 2007; Dimri et al., 2002; Haupt et al., 1991; Leung et al., 2004; van Lohuizen et al., 1991; Vonlanthen et al., 2001). Other PcG proteins have since also been implicated in tumourigenesis. Increased levels of CBX7 have been found to be involved in lymphomas and prostate cancer (Bernard et al., 2005). Whereas EZH2 is commonly found overexpressed in various types of cancer, including leukaemia, prostate cancer and breast cancer and is regarded as a potential marker of advanced aggressive cancer with poor prognosis (Bracken et al., 2003; Kleer et al., 2003; Varambally et al., 2002). SUZ12 is also found to be upregulated in various cancers, including colon, breast and liver cancer (Kirmizis et al., 2003).

The dual role of PcG proteins in cancer and stem cells is unlikely to be coincidental. Over recent years there has been emerging evidence that tumours contain a subset of cancer cells that display numerous properties of stem cells, particularly the ability to self-renew and to spawn differentiated progeny. Pioneering studies on leukaemia demonstrated the existence of a small subset of leukaemic cells that carry specific cell surface markers and are able to seed new tumors (Lapidot et al., 1994). Further studies identified similar cells from solid breast and brain tumours (Al-Hajj et al., 2003; Singh et al., 2003). These findings have provided the basis for the cancer stem cell hypothesis, which holds that the origin of tumours are based on stem cell properties, either as a basis (somatic stem cells or progenitor cells) or acquired (terminally differentiated cells) in the process of transformation. The increased understanding of what governs the plasticity of cellular identity and the transcriptional events to achieve pluripotency, in the form of iPS cells, has shed light on the processes involved in reversing differentiation and acquiring self-renewal capacity. The role of PcG proteins in resetting the repression of differentiation-specific genes is likely to be an important topic ((Endoh et al., 2008), reviewed in (Jaenisch and Young, 2008)). Therefore, the

role of PcG proteins in stem cells and differentiation and the possible contribution to tumourigenesis can be proposed as shown in Figure 1.2.

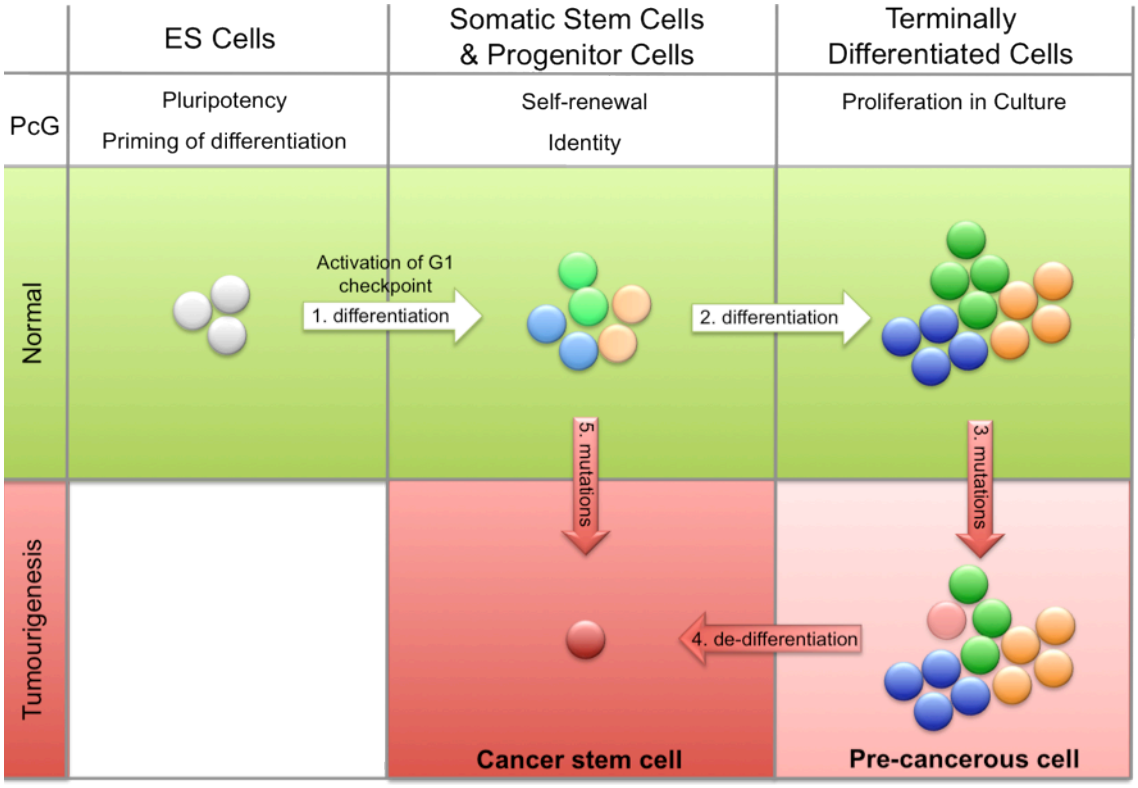


Figure 1.2 Model for PcG involvement in normal differentiation and opportunities for tumourigenesis

PcG repression is involved at various stages of differentiation. In ES cells (white circles) PcG complexes are responsible for the maintenance of pluripotency by repression of differentiation-specific genes. PcG binding at these genes creates a bivalent state, from which either activation or silencing can be quickly achieved upon differentation. 1) differentiation to a multipotent or oligopotent somatic stem- or progenitor cell (light green, blue and orange circles) results in the loss of pluripotency, which is mediated by loss of PcG repression at differentiation-related genes. Differentiation also activates more profound cell cycle control, including the activation of the G1 checkpoint, and PcG repression is needed to repress the genes limiting proliferation in order to maintain the self-renewal capacity. 2) when the cell becomes terminally differentiated (dark green, blue and orange circles), self-renewal is lost and the cells become quiescent, unless stimulated in culture. 3) however, due to the accumulation of mutations over time a differentiated cell may acquire properties that will allow it to become susceptible to transformation and therefore a pre-cancerous cell

(light red circle). 4) complete tumourigenesis would require unlimited proliferation and therefore could involve de-differentiation to acquire the self-renewal properties of adult stem-/progenitor cells to become a cancer stem cell (red circle). 5) alternatively, somatic stem- or progenitor cells could directly acquire the transforming mutations to become a cancer stem cell.

1.3 PcG proteins in regulation of the INK4A tumour suppressor

So far, the focus has been on the role of PcG-mediated repression for the control of developmentally regulated genes. However, one of the critical targets identified in the *Bmi1* knockout mice was the *Ink4a/Arf* locus (Jacobs et al., 1999a). This section aims to introduce the products of the *INK4B-ARF-INK4A* locus, their known properties and functions, and their involvement in the implementation of senescence, the state of permanent growth arrest activated in response to various types of cellular stress, which represents an important barrier in tumourigenesis and sustained somatic stem cell self-renewal.

1.3.1 The *INK4B-ARF-INK4A* locus and its products

The *INK4B-ARF-INK4A* locus on human chromosome 9p21, and the corresponding locus on mouse chromosome 4, comprises two annotated genes, officially termed *CDKN2B* and *CDKN2A*, but more commonly referred to as *INK4B* and *INK4/ARF* respectively. These genes encode three proteins, commonly referred to as $p15^{INK4B}$, $p14^{ARF}$ ($p19^{Arf}$ in mice) and $p16^{INK4A}$ (see Figure 1.3). The $p15^{INK4B}$ and $p16^{INK4A}$ proteins are cyclin-dependent kinase (CDK) inhibitors that bind directly to CDK4 and CDK6 and thereby prevent their interaction with regulatory D-type cyclins. Inhibition of CDK4 and CDK6 blocks the hyperphosphorylation of the RB tumour suppressor protein. In its hypophophorylated and active form, RB functions as a negative regulator of the E2F transcription factors (reviewed in (Classon and Harlow, 2002)). E2F activates a series of genes involved in DNA replication and nucleotide metabolism that are required for cell cycle progression into S-phase (Dyson, 1998).

The $p16^{INK4A}$ and $p15^{INK4B}$ proteins are closely related (85% similarity in primary sequence) and it is believed that *INK4A* arose in evolution via gene duplication of

INK4B (Gil and Peters, 2006; Kim and Sharpless, 2006). In contrast, the ARF protein has no structural or functional counterpart conserved throughout evolution. It is encoded by an exon, designated exon 1β, that is located between *INK4A* and *INK4B* and becomes spliced to exon 2 of *INK4A*. The acronym ARF reflects the fact that it incorporates amino acids encoded by the shared exon in an alternative reading frame relative to *INK4A* (Quelle et al., 1995). ARF is thought to interact with a wide variety of cellular proteins. Its best understood property is that it can stabilise and activate the p53 tumour suppressor protein by inhibiting its ubiquitination by HDM2 (reviewed in (Vousden and Prives, 2009)).

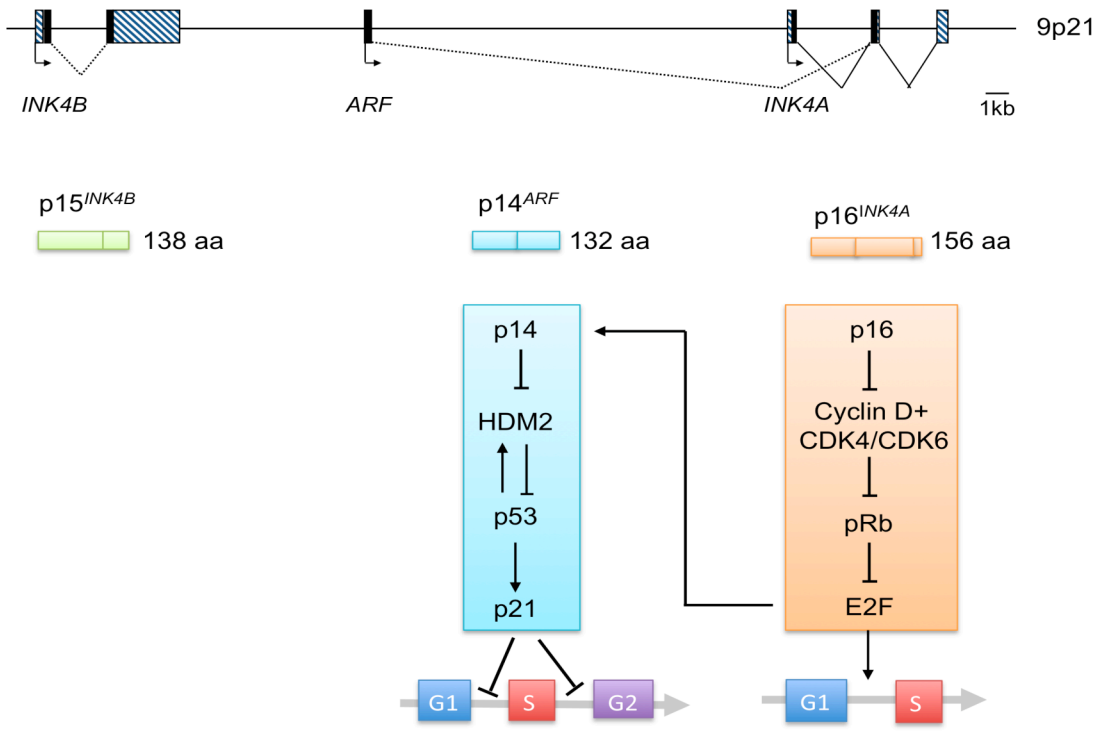


Figure 1.3 The human *INK4B-ARF-INK4A* locus and its products

The *INK4/ARF* locus encodes two genes *CDKN2A* and *CDKN2B* that encode three coding transcripts; two highly related genes *INK4A* and *INK4B*, and an unrelated gene *ARF*. The product of the latter, $p14^{ARF}$ protein is a potent inhibitor of HDM2, which in its turn is an inhibitor the p53 tumour suppressor protein. Stabilisation of p53 leads to the transactivation of a wide array of genes, including $p21^{CIP1}$ that can arrest the cell cycle at multiple stages by inhibiting the respective cyclin-dependent kinases (CDKs). $P16^{INK4A}$ also inhibits cell cycle progression at G1/S phase through the specific inhibition of CDK4 and CDK6. This blocks the phosphorylation of pRb and allows it to inhibit the E2F transcription factors, which are essential for cell cycle progression to S phase. $P15^{INK4B}$ works through the same mechanism as p $P16^{INK4A}$.

One of the most important downstream targets of activated p53 is the CDK inhibitor $p21^{CIP1}$, which is crucial for p53-dependent cell cycle arrest in response to DNA damage (Brown et al., 1997). Interestingly, ARF also forms a connection between the p53- and RB pathways as it is transcriptionally activated by E2F (Bates et al., 1998). Thus, agents, such as viral oncoproteins, that inactivate RB will inevitably invoke a p53 response via ARF.

1.3.2 Role of the *INK4A-ARF-INK4B* locus in senescence

It is evident that by activating the RB and p53 pathways, all three of the proteins encoded by the *INK4B-ARF-INK4A* locus have the ability to cause cell cycle arrest. However, there is no evidence that they are involved in the regulation of the normal cell cycle. Their levels do not fluctuate significantly as cells progress through different phases and at least in the case of the *INK4* genes, the encoded proteins are extremely stable. They also appear to be expressed at very low, essentially undetectable levels in most normal tissues. The one process however in which all three proteins appear to have an important role is the implementation of senescence.

The term "senescence", derived from *senex*, Latin for old man or old age, refers to a state of irreversible growth arrest that is now recognised as a hallmark of tumour suppression and ageing. The concept of senescence was originally proposed by Hayflick and colleagues who found that primary human diploid fibroblasts (HDFs) grown in tissue culture achieve a specific but finite number of population doublings (Hayflick and Moorhead, 1961). In the following decades, these observations have been developed and extended to other cell types but HDFs remain the system of choice for experimental studies on senescence. Importantly, when the cells reach the end of their finite lifespan, they adopt a characteristic phenotype that distinguishes them from cells arresting in response to growth factor withdrawal or contact inhibition. These include an enlarged and flattened appearance, failure to respond to proliferative stimuli, and resistance to apoptosis. They also develop single prominent nucleoli, senescence-associated heterechromatin foci (SAHF) and -DNA damage foci (SDF), and stain positively for senescence-associated β-galactosidase activity (SA-βgal) (reviewed in (Campisi and d'Adda di Fagagna, 2007)).

Another important feature of senescent cells is altered patterns of gene expression, including the upregulation of secretory proteins involved in changing the

extracellular matrix or in local inflammation, such as IL6, IL8 and the IL8 receptor CXCR2 (Acosta et al., 2008; Kuilman et al., 2008; Rodier et al., 2009). However, from the perspective of this thesis, the most noteworthy aspect of the senescence phenotype is the near uniform increase in expression of the $p16^{INK4A}$ and $p21^{CIP1}$ CDK inhibitors and various lines of evidence have confirmed that these proteins are instrumental in bringing cell proliferation to a halt (reviewed in (Besson et al., 2008; Campisi and d'Adda di Fagagna, 2007)).

1.3.3 Different Types of Senescence

The classical studies on senescence in human fibroblasts suggested that there must be an intrinsic counting mechanism that pre-determines the lifespan of particular cell strains. The accepted explanation is that this counting mechanism is provided by the progressive erosion of telomeric repeats that occurs with each cell division, due to the inability of the replication machinery to copy the ends of linear DNA molecules (reviewed in (Deng et al., 2008)). Complete loss or uncapping of the telomeric DNA produces the equivalent of a double strand break, which triggers the DNA damage response (DDR) and activation of p53 (d'Adda di Fagagna et al., 2003). This form of senescence, commonly referred to as replicative senescence, can be prevented by the activation of telomerase, the enzyme responsible for maintaining telomere length. In general, this enzyme is either not expressed or present at insufficient levels in normal somatic cells (Bodnar et al., 1998; Masutomi et al., 2003). (Blasco et al., 1997; Masutomi et al., 2003). However, telomerase is frequently activated in cancer cells, enabling them to avoid senescence (reviewed in (Stewart and Weinberg, 2006)).

Importantly, there are other scenarios in which cells develop the characteristics of senescence, irrespective of telomere status. Two typical examples are human epithelial cells, which undergo a senescence-like arrest in tissue culture after only a few population doublings, and mouse embryonic fibroblasts (MEFs), which have a limited lifespan despite having exceptionally long telomeres. The explanation is that standard conditions of tissue culture can be sufficiently stressful to engage telomere-independent senescence, sometimes referred to as ‘culture shock’. The effects of culture shock can be alleviated by growing cells on feeder layers (Ramirez et al., 2001) or, in the case of MEFs, by growing the cells in reduced oxygen concentrations that are closer to physiological levels (Parrinello et al., 2003). This is thought to reduce the oxidative

DNA damage caused by reactive oxygen species (ROS) and thereby reduce the chances of activating the DDR and engaging senescence (reviewed in (Lu and Finkel, 2008)).

From the perspective of tumour suppression, the most significant type of stress that can elicit a senescence response is aberrant signalling from an activated oncogene. Oncogene-induced senescence (OIS) was first reported in primary fibroblasts exposed to oncogenic RAS (RAS^{V12}) and was shown to be associated with elevated levels of $p16^{INK4A}$ and p53 (Serrano et al., 1997). Similar findings were recorded with activated RAF and MEK, downstream components of the RAS signalling pathway (Lin et al., 1998; Zhu et al., 1998). The mechanisms underlying OIS and how they impact on the *INK4B-ARF-INK4A* locus remain poorly understood but one common denominator could be the DDR that results from increased DNA replication caused by these aberrant mitogenic stimuli (Bartkova et al., 2006; Di Micco et al., 2006).

1.3.4 *INK4B-ARF-INK4A,* senescence and tumour suppression

The involvement of *INK4A* in the implementation of OIS would provide an attractive explanation for its role as a tumour suppressor. It was first identified as such from studies showing that predisposition to melanoma could in certain families be linked to markers on chromosome 9p21 and ultimately to inactivating mutations in *INK4A* (Kamb et al., 1994). It has subsequently been shown that *INK4A* is frequently mutated, deleted or epigenetically silenced in a large variety of primary cancers and tumour cell lines (reviewed in (Ruas and Peters, 1998; Sharpless, 2005)). Indeed, it has been claimed that $p16^{INK4A}$ is second only to p53 in the frequency with which it is inactivated in human cancers. However, many of the homozygous deletions that occur on chromosome 9p21 encompass *ARF* and *INK4B* making it difficult to determine what contributions these two genes might have in tumour suppression. Missense mutations in *ARF* and *INK4B* are extremely rare but there are circumstances in which they appear to have been selectively inactivated by deletion or methylation (Peters, 2008).

Curiously, a very different picture has emerged from studies on the mouse locus. Mice that are defective for *Arf* specifically are highly tumour prone and the corresponding MEFs are immortal and sensitive to transformation by RAS (Kamijo et al., 1997). In contrast, mice that specifically lack *Ink4a* have a relatively low incidence of spontaneous tumours, unless challenged with a chemical carcinogen, and the *Ink4a* -/- MEFs undergo senescence (Krimpenfort et al., 2001; Sharpless et al., 2001). Thus, all

the evidence suggests that the *Arf/p53* pathways is critical for tumour suppression in mice. In line with this, RAS activates the expression of mouse *Arf* but has little effect on human *ARF*. Studies in mice have also confirmed a role for $p15^{Ink4b}$ in tumour suppression. Although the single knockout of *Ink4b* displays only very subtle tumour predisposition, the combinatorial deletion of all three genes at the locus implies that loss of *Ink4b* significantly exacerbates the tumour prone phenotype of the *Ink4a/Arf* knockout (Krimpenfort et al., 2007). It is noteworthy that $p15^{INK4B}$ was identified as one of the most robust markers for OIS in human fibroblasts exposed to oncogene RAS (Collado et al., 2005).

Importantly, several related studies demonstrated that OIS is not confined to the artificial conditions of tissue culture but is clearly observed in benign and pre-malignant tumours in vivo, both in mouse models and in humans (Braig et al., 2005; Chen et al., 2005; Collado et al., 2005; Michaloglou et al., 2005). One of the most compelling examples is the accumulation of $p16^{INK4A}$ and SA-βgal in benign melanocytic naevi that have an activated allele of B-RAF (Michaloglou et al., 2005).

1.3.5 Senescence and ageing

As an intrinsic limit on the proliferative capacity of cells, senescence is on the one hand seen as an important tumour suppressive mechanism, limiting unscheduled proliferation and the accumulation of mutations. On the other hand, limited proliferation of mitotic cells, especially somatic stem- and progenitor cells, will limit the self-renewal ability of tissues and thereby contribute to organismal ageing. In evolutionary terms, these opposing roles are referred to as antagonistic pleiotropy. Senescence has evolved from a time when longevity was highly constrained by a multitude of environmental factors, such as bacterial or viral infection, starvation and predation, and hence the negative contributions of senescence to ageing was never subject to strong selective pressure.

Several studies have reported increased numbers of senescent cells in ageing tissue, as judged by SA-βgal (Dimri et al., 1995) and $p16^{INK4A}$ staining (Krishnamurthy et al., 2004; Ressler et al., 2006). It has been suggested that the secretory programme of senescent cells could have an impact on the wider tissue environment and thereby affect age-related decrements in tissue structure and function (Kuilman and Peeper, 2009). Furthermore, senescence occurs at sites of age-related disorders, including osteoarthtitis and artherosclerosis (reviewed in (Campisi and d'Adda di Fagagna, 2007). Recent

association studies have identified genetic variants near the *INK4/ARF* locus linked to age-related disorders, including type II diabetes, coronary artery disease and general frailty (Helgadottir et al., 2007; McPherson et al., 2007; Saxena et al., 2007; Scott et al., 2007b; Zeggini et al., 2007). Although it is not known how these variants affect the expression of the *INK4/ARF* locus, the putative contribution to the ageing phenotype provides an interesting link.

As well as these correlative studies, there is experimental evidence linking increased $p16^{INK4A}$ levels with age-related disorders in mice engineered to have an extra copy of the *Ink4a/Arf* locus (Krishnamurthy et al., 2006; Matheu et al., 2004). Although these mice did not show any decrease in longevity, the self-renewal capacity of pancreatic β islet cells was clearly impaired, leading to diabetes. These findings are consistent with earlier observations that *Cdk4* null mice are diabetic. Conversely, deletion of *Ink4a* delays the age-related loss of islet cell plasticity although these mice were more prone to developing tumours. In addition, p16-independent loss of self-renewal has been observed, suggesting that there must be other contributors to ageing, with ARF and its action on the p53 pathway being an attractive candidate (reviewed in (Kim and Sharpless, 2006).

It is interesting to speculate how and when the repression of *INK4A/ARF* needs to be established in embryonic development. Due to the high rate of proliferation needed for the expanding embryo, ES cells have been found to have a significantly shortened cell cycle without a classical G1 checkpoint (Orford and Scadden, 2008). This is illustrated by high levels of Cyclin D3 and pRbm and the irresponsiveness of Cyclin D3-Cdk6 complexes to ectopic expression of *Ink4a* (Faast et al., 2004). In addition, genome-wide binding assays of PcG proteins in ES cells did not detect binding to *Ink4a/Arf* ((Boyer et al., 2006)). It could therefore be that the repression of *INK4A/ARF* only becomes relevant at later stages in development when other, p16-sensitive Cdk-Cyclin D complexes become active. Repression of *INK4A/ARF* may therefore be a common part of the differentiation program initiated through PcG. Interestingly, the activation of $p16^{INK4A}$ and $p21^{CIP1}$ acts as a barrier in the reprogramming of somatic cells to become induced pluripotent stem (iPS) cells (Banito et al., 2009; Li et al., 2009; Marion et al., 2009; Mikkelsen et al., 2008). This would suggest that PcG-mediated repression of *INK4A/ARF* is an important event in the acquisition of pluripotency.

1.3.6 PcG-mediated regulation of *INK4A*

As alluded to earlier, studies on *Bmi1* null mice provided the first evidence for a link between PcG proteins and the regulation of *Ink4a/Arf* and a clear inference was that the impaired proliferation of *Bmi1*-/- MEFs reflected derepression of *Ink4a/Arf* and premature senescence (Jacobs et al., 1999a). Conversely, overexpression of Bmi1 could downregulate expression of $p16^{Ink4a}$ and $p19^{Arf}$ in MEFs, leading to immortalization (Jacobs et al., 1999a). These ideas were subsequently confirmed and extended by a variety of genetic crosses that showed that features of the *Bmi*1 null phenotype, including the neurological and haematological stem cell defects, can be rescued by simultaneous ablation of *Ink4a/Arf* (Molofsky et al., 2003; Park et al., 2003). Conversely, it was found that up-regulated *Bmi1* expression leads to enhanced β islet regeneration by enhanced repression of *Ink4a* (Dhawan et al., 2009). The further use of mice that are selectively defective for either *Ink4a* or *Arf* suggested a role for both these genes in specific subsets of stem and progenitor cells (Janzen et al., 2006; Lewis et al., 2001; Molofsky et al., 2006). However, analogous experiments in human fibroblasts using shRNA-mediated knockdown or retrovirally transduced cDNA suggests that the effects of Bmi1 are limited to *INK4A* and do not extend to ARF (Bracken et al., 2007; Maertens et al., 2009). In the remainder of this thesis, the emphasis will be on the effects of PcG proteins on the human *INK4A* locus.

As Bmi1 is believed to act as part of the multicomponent PRC1 complex, it would be logical to expect other PRC1 proteins to have a similar influence on *INK4A*. Studies of a number of PcG gene knockouts in mice indeed reported that some of the defects associated with ablation of *Mel18, Mph2*, *Ring1B* were, at least in part, attributable to derepression of *Ink4a* (Cales et al., 2008; Isono et al., 2005a; Jacobs et al., 1999a; Miki et al., 2007; Voncken et al., 2003). Similar implications were drawn for *Ezh2* (Chen et al., 2009; Krishnamurthy et al., 2006), but not for PcG genes *Eed* and *Mph1* (Lessard et al., 1999; Ohta et al., 2002). Extending these studies to human cells, it has been shown that shRNA-mediated knockdown of CBX7, CBX8, MEL18 and BMI1 results in elevated expression of *INK4A* and premature senescence (Bracken et al., 2007; Dietrich et al., 2007; Gil et al., 2004; Jacobs et al., 1999a; Maertens et al., 2009) whereas overexpression of the mouse homologues Cbx7, Cbx8 and Bmi1, used for technical reasons, in human cells can repress *INK4A* and delay senescence (Dietrich et al., 2007; Gil et al., 2004; He et al., 2009).

Given these observations, it is evident that the role of PcG proteins in cancer can be largely attributed to the repression of *INK4A*. For example, EZH2, which is overexpressed in a variety of tumours, was shown to repress *INK4A* directly and to be responsible for additional repression by *de novo* methylation of *INK4A* in intrahepatic cholangiocarcinoma (Bracken et al., 2003; Sasaki et al., 2008). Interestingly, a number of cancer types that overexpress PcG proteins were found to have an intact *INK4A* locus (Bernard et al., 2005; Scott et al., 2007a), suggesting that PcG-mediated represson of *INK4A* provides an alternative to deletion or mutation of the locus as a step to bypass senescence (reviewed in (Valk-Lingbeek et al., 2004)).

However, as PcG proteins repress a large number of target genes in addition to *INK4A*, it is of course feasible that they promote tumorigenesis in other ways. Indeed it has been shown that Bmi1 can engage EGFR-induced transformation (Bruggeman et al., 2007; Kang et al., 2007; Liu et al., 2006) and that the enhanced self-renewal of haematopoietic stem cells (HSCs) caused by Bmi1 overexpression is manifested in the absence of *Ink4a/Arf*. Moreover, gene expression profiling of *Bmi1*$^{-/-}$ HSCs found increased expression of stem cell associated genes, cell survival genes and transcription factors as well as *Ink4a/Arf* (Iwama et al., 2004; Park et al., 2003). Also, the key developmental signalling pathways operated by the SHH, WNT and NOTCH proteins are increasingly implicated in PcG regulation in cancer (reviewed in (Sparmann and van Lohuizen, 2006).

Recently, it has also been reported that the induction of *INK4A* (and *Ink4a/Arf* in mice) by oncogenic RAS involves the JMJD3 histone demethylase (Agger et al., 2009; Barradas et al., 2009). RAS was found to downregulate EZH2 and upregulate JMJD3, which facilitates the demethylation of the H3K27me3 mark and removal of PcG complexes from the locus. Knockdown of JMJD3 was found to delay senescence and immortalise MEFs. Taken together, these findings suggest that JMJD3 is a clear antagonist of PcG-mediated repression of *INK4A* and might act as a tumour suppressor.

1.4 Characterising PRC1 complexes that regulate INK4a

At the start of this work, it was already known that several core components of the PRC1 complex are implicated in the regulation of *INK4A*, as described in the previous section. Importantly, these included two members of the Pc family, CBX7 and CBX8, and two Psc proteins, BMI1 and MEL18. There were also indications from the analyses of affinity-purified complexes, that the mammalian PRC1 complexes contain single representatives of the Pc, Psc, Ph and Sce families, in line with the *Drosophila* paradigm. In principle, therefore, the mammalian PcG proteins described in Table 1.1 have the potential to form 180 different permutations of PRC1 complex.

One way to try to resolve some of these issues would be conduct a comprehensive survey to identify all of the PcG proteins that are involved in the regulation of *INK4A*. However, as a starting point, it was decided to focus on the HPH proteins for a number of reasons. First, they are arguably the least well characterised of the PRC1 proteins in terms of their structural or functional contribution to the complex. Second, of the three known Ph proteins in mammalian cells, only one has thus far been shown to influence expression of *Ink4a* in mice, namely Mph2, whereas family member Mph1 did not (Isono et al., 2005a; Ohta et al., 2002). Finally, preliminary findings in the laboratory identified HPH2 as the only member of the family that co-purified with mouse Cbx7 that is known to regulate *INK4A*.

1.4.1 HPH Family overview

As for all other PcG genes, *polyhomeotic* was initially discovered in mutant screens in *Drosophila*, where there are two nearly identical copies, Ph proximal and Ph distal (Dura et al., 1985; Dura et al., 1987). Early analyses already provided clues about functional domains and suggested a putative role as a transcription factor, as part of a protein complex (Deatrick et al., 1991). As these domains have been conserved in evolution, they provide a basis for seeking functions common to all members of the family (Nomura et al., 1994). The mammalian polyhomeotic family is composed of three genes, which products are smaller in size than the *Drosophila* protein (Figure 1.4). In addition, Mph2 has also been reported to have a short isoform (Yamaki et al., 2002). Consistent with the *Drosophila* genes, the conserved domains are located in the carboxy-terminal third of each protein. These are designated homology domain 1

(HD1), the (Cys)4-type Zn coordination- or FCS domain, and the homology domain II (HD2) or sterile alpha motif (SAM) domain. Since, the conserved domains only comprise a small part of each protein and as the PH proteins are considerably larger than other PRC1 proteins there would in principle be ample opportunity for additional, perhaps unique functions for each individual protein.

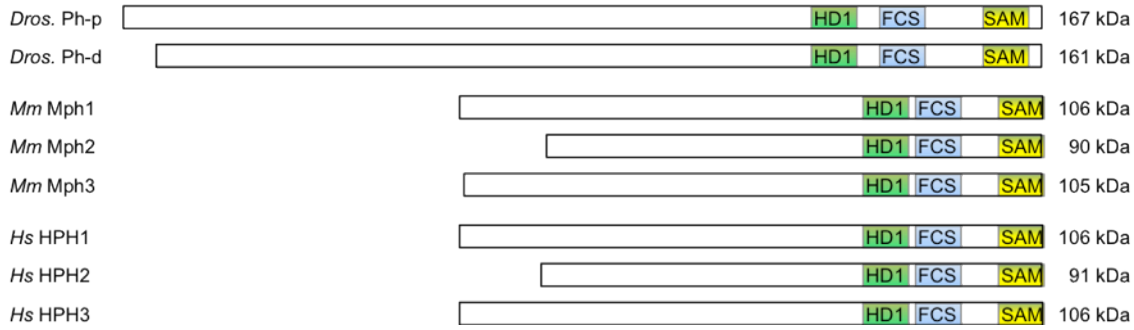


Figure 1.4 Conservation of polyhomeotic family in different species
Schematic representation of the polyhomeotic proteins in Drosophila melanogaster, Mus musculus and Homo sapiens, indicating their molecular weights and the relative locations of the HD1, FCS and SAM converved C-terminal functional domains (as derived from the Uniprot database).

1.4.1.1 SAM Domain Oligomerisation

The sterile alpha motif (SAM) domain has been shown to mediate binding between two polyhomeotic proteins, even when originating from two distinct species (Kyba and Brock, 1998b). It was found that the resulting oligomer of purified *Drosophila* Ph SAM domains forms an alpha helical structure, with 6 subunits per turn (Kim et al., 2002). In addition to homotypic SAM interactions between polyhomeotic proteins, heterotypic SAM interactions with the non-core polycomb protein SCM have been observed, again forming a helical polymer (Kim et al., 2005; Kyba and Brock, 1998b; Peterson et al., 1997). In addition, binding studies with HPH1 and HPH2 reported that the HD1 domain of one protein interacts with the SAM domain of the other (Gunster et al., 1997).

The self-association of the different Phs could provide a mechanism for the spreading of Polycomb complexes along the chromatin of their target genes or to bridge between loci from different intra- or interchromosomal locations. In addition, it was suggested that the structure of the chromatin fiber wrapped around the helical polymer

of SAM domains would be thermodynamically stablilised by protein-DNA and protein-histone interactions (Kim et al., 2002). Although there is *in vitro* binding data that suggests Psc and Pc can also self-associate under certain conditions (Kyba and Brock, 1998a; Min et al., 2003), oligomerisation properties of the polyhomeotics make it a strong suspect for chromatin spreading of PRC1 complexes,

1.4.1.2 Association with MAPKAP and ATM/ATR Kinases

The polyhomeotic proteins have been shown to interact with the mitogen activated protein kinases-associated protein kinases (MAPKAP or MK). MKs function downstream of the ERK pathway following mitogen stimulation and the p38 MAPK cascade in response to stress signals. MK2, MK3 and MK5 are involved in multiple cellular functions, including cell-cycle control, cellular differentiation and chromatin repression and remodelling (reviewed in (Gaestel, 2006)). MK2 was found to interact with HPH2 in the context of PRC1 and to co-localise to Polycomb bodies (Schwermann et al., 2009; Yannoni et al., 2004). HPH2 binding of MK2 was required to rescue the reduced self-renewal capacity of $MK2^{-/-}$ HSCs, associated with increased levels of $p19^{Arf}$, by ectopic MK2 (Schwermann et al., 2009). This study also reported *in vitro* interactions between HPH1 and MK2, and between HPH2 and MK3, but not MK5. HPH1 and BMI1 were also identified as substrates for MK3 phosphorylation and MK3 overexpression results in the release of PRC1 proteins from chromatin and reduced proliferation accompanied by a loss of PcG-mediated repression of $p14^{ARF}$ ($p16^{INK4A}$ was not tested) (Voncken et al., 2005). This however opposes the findings for loss of MK2 in HSCs or other phosphorylation events of BMI1 needed for its chromatin association (Schwermann et al., 2009; Voncken et al., 1999).

In addition, HPH2 and HPH3 were found to be substrates for the DNA damage response kinases ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) in a large proteomic screen in 293T cells (Matsuoka et al., 2007). Interestingly, the phosphorylation sites, S733 in HPH2 and S229 in HPH3, are in largely distinct locations and do not correspond to conserved motifs among the HPH proteins. Upon closer inspection, however, there are multiple PQ and SQ motifs in HPH1 and HPH2 in the regions aligning with S229 of HPH3. Additionally, the Swissprot database lists three other predicted phosphoserines in the vicinity of S733 in HPH2.

1.4.1.3 Nucleic Acid Binding

As previously mentioned, there is a compelling link between RNA and PcG-mediated repression but it is not clear which PcG proteins are involved in the interaction with RNA. Although the chromodomains of the CBX and SCM proteins have been shown to interact with RNA (Akhtar et al., 2000; Bernstein et al., 2006b), there is also evidence ascribing nucleic acid interactions to the functional repertoire of the polyhomeotics. *In vitro* RNA binding studies with Mph1 identified the conserved zinc finger domain ($CX_2C...CX_3C$ type), often referred to as the FCS finger, as the region of interaction with both RNA and DNA (Zhang et al., 2004). The binding was found not to be sequence-specific and both single- and double-stranded RNAs (but not tRNA) were bound by the domain. A recent structure of the FCS domain from L3(MBT)-like shows an RNA binding pocket formed by the conserved cysteines (Lechtenberg et al., 2009) and mutation of the cysteine residues perturbs RNA binding. Similar observations apply to the *C. elegans* protein SOP2, which is a distant relative of Ph and Scm. SOP2 RNA binding mutants were shown to display homeotic transformations and loss of nuclear localisation, suggesting an essential role for RNA in guiding Polycomb-mediated repression (Zhang et al., 2004; Zhang et al., 2006).

1.4.1.4 Interaction with Transcription Factors

Whereas genetic ablation of *Mph2* leads to impaired proliferation, a recent report has suggested that loss of heterozygosity (LOH) of HPH3, which is frequently observed in osteosarcoma, contributes to tumorigenesis by disrupting quiescence (Deshpande et al., 2007). HPH3 has been shown to interact with E2F6, which is believed to induce a G_0 arrest by acting as a dominant negative regulator of the E2F family of transcription factors (Ogawa et al., 2002). Along similar lines, E2F6 had previously been shown to interact with the PcG proteins RING1A, RING1B, MEL18 in human, and Bmi1 and Mph1 in mouse, at E2F- and Myc-responsive proteins (Deshpande et al., 2007; Ogawa et al., 2002; Trimarchi et al., 2001). It is conceivable, therefore, that loss of HPH1 or HPH3 and associated PcG function could lead to derepression of the E2F- and Myc-responsive genes.

On the contrary, HPH2 was found to interact with the activation domains of the transcription factor Sp1 in a yeast two-hybrid screen (Gunther et al., 2000). Sp1 is

widely involved in transcriptional regulation and interacts both with general transcription factors that constitute the transcription machinery and a wide array of DNA binding regulatory factors, including RB, p107, p53, E2F and YY1, and the basal transcription machinery (reviewed in (Wierstra, 2008)). Furthermore, in Drosophila Sp1 has been found at PREs and therefore might be a co-factor for PcG recruitment to target genes (Brown et al., 2005). The interaction between Sp1 and HPH2 was validated by in vitro binding and involved the first 266 amino acids (Gunther et al., 2000). This is of interest, as this region does not include any of the conserved domains of polyhomeotics, suggesting that this interaction could indeed be unique to HPH2.

In summary, the HPH proteins thus present a number of interesting properties and interactions in the context of PcG-mediated repression, some of which present a putative connection to the wider domain of *INK4A* function.

1.5 Thesis Aims

The general aim of this thesis was to investigate the potential role of the HPH proteins in the regulation of *INK4A* and more specifically to address the following questions:

1. Is there any evidence for specificity in the interaction between HPH2 and other PRC1 components that can be used as a foundation for defining the PRC1 complexes that regulate *INK4A*?

2. Is HPH2 and/or other members of the HPH family directly involved in the regulation of *INK4A*?

3. What functions does HPH2 contribute to the regulation of *INK4A*?

2 Materials and Methods

2.1 Solutions

Versene, PBS-A, LB, LB-Agar and clean glassware were provided by the LRI Media Services. Chemicals were generally purchased from Sigma-Aldrich. Water was purified using a Millipore reverse osmosis system. Solutions were made up in water (unless stated otherwise) and either sterilised through autoclaving or filtered through a 0.22 μm filter unit (Millipore) if needed.

General solutions:

PBS-A	8.06 mM Na_2HPO_4
	0.8% (w/v) NaCl
	1.47 mM KH_2PO_4
	0.025% (w/v) KCl (pH 7.2)

Bacterial Culture & Lysis

Luria broth (LB)	1% (w/v) NaCl
	0.5% yeast extract
	1% (w/v) bacto-tryptone
SOC medium	2% (w/v) tryptone
	0.05% (w/v) yeast extract
	0.006% (w/v) NaCl
	0.002% (w/v) KCl
	0.002% (w/v) $MgCl_2$
	0.036% (w/v) D-glucose
GST Lysis buffer	10 mM TRIS HCl pH 7.6
	1 M NaCl
	0.1 mM PMSF

GST elution buffer	10 mM reduced glutathione 50 mM TRIS HCl pH 8.0
HIS purification buffer	50 mM NaH_2PO_4 pH 8.0 0.5 M NaCl
HIS imidazole buffer	3 M imidazole 0.5 M NaCl 20 mM NaH_2PO_4 pH 6.0

Cell Culture:

2x HBS	200 mM NaCl 1.5 mM Na_2PO4 50 mM HEPES (pH 7.1)
Versene	0.02% (w/v) EDTA in PBS-A 0.0015% (w/v) phenol red

SDS-PAGE and Western blotting:

RIPA lysis buffer	25 mM TRIS HCl pH 7.6 150 mM NaCl 1% (v/v) NP40 1% (v/v) sodium deoxycholate 0.1% (v/v) SDS 0.1 mM PMSF 1x Protease Inhibitor Cocktail (Sigma)
Protein running buffer	25 mM TRIS base 192 mM glycine 0.1% (w/v) SDS
Protein transfer buffer	Protein running buffer

	20% (v/v) methanol
5x Laemmli buffer	0.375 M TRIS HCl pH 6.8 10% (v/v) SDS 50% (v/v) glycerol 0.01% (w/v) bromophenol blue 25% (v/v) 2-mercaptoethanol
WB washing buffer	PBS-A 0.1% (v/v) Tween-20
WB blocking buffer	WB washing buffer 5% dried milk (Marvel)
Western stripping buffer	0.5 M TRIS HCl pH 6.8 20% (v/v) SDS 0.5% (v/v) 2-mercaptoethanol
Fixing/Destain solution	50% (v/v) methanol 10% (v/v) acetic acid
Coomassie staining solution	0.1% (w/v) Coomassie blue R250 50% (v/v) methanol 10% (v/v) acetic acid

<u>Immunoprecipitation</u>

NP40 lysis buffer	50 mM TRIS HCl pH 8.0 150 mM NaCl 1 mM EDTA 1% (v/v) NP40 0.1 mM PMSF 1x Complete EDTA free (Roche), or; 1x Protease Inhibitor Cocktail (Sigma)

ChIP buffer	50 mM TRIS HCl pH 7.5
	150 mM NaCl
	5 mM EDTA
	1% (v/v) NP40
	1% (v/v) Triton X-100
	0.1 mM PMSF
	1x Complete EDTA free (Roche)
ChIP elution buffer	50 mM $NaHCO_3$
	1% (v/v) SDS

RNA ChIP[1]

Buffer A *(with or without NP40)*	5 mM PIPES pH 8.0
	85 mM KCl
	1X Complete EDTA free (Roche)
	50 U/ml SUPERase in (Ambion)
	0.5% (v/v) NP40
Buffer B	1% (v/v) SDS
	10 mM EDTA
	50 mM TRIS HCl pH 8.1
	1X Complete EDTA free (Roche)
	50 U/ml SUPERase in (Ambion)
IP buffer	0.01% (v/v) SDS
	1.1% (v/v) Triton X-100
	1.2 mM EDTA
	16.7 mM TRIS HCl pH 8.1
	167 mM NaCl
	1X Complete EDTA free (Roche)
	50 U/ml SUPERase in (Ambion)

[1] www.epigenetics-noe.net/researchtools/protocol.php?/protid=28reagents

Low-salt wash buffer	0.1% (v/v) SDS 1% (v/v) Triton X-100 2 mM EDTA 20 mM TRIS HCl pH 8.1 150 mM NaCl
Low-salt wash buffer	0.1% (v/v) SDS 1% (v/v) Triton X-100 2 mM EDTA 20 mM TRIS HCl pH 8.1 500 mM NaCl
LiCl wash	0.25 M LiCl 1% (v/v) NP40 1% (v/v) deoxycholate 1 mM EDTA 10 mM TRIS HCl pH 8.1
Elution buffer	1% SDS 0.1 M $NaHCO_3$ 50 U/ml SUPERase in (Ambion)

<u>DNA analysis:</u>

Primer annealing buffer	100 mM K-acetate 30 mM hepes-KOH pH 7.4 2 mM Mg-acetate
DNA loading buffer	60% (v/v) sucrose 0.1% (w/v) bromophenol blue
DNA running buffer	40 mM TRIS-acetate (pH 7.6) 5 mM sodium acetate

	1 mM EDTA

Tandem Affinity Purification:

Lysis buffer	50 mM TRIS HCl pH 8.0
	0.3 M NaCl
	1 mM EDTA
	1.02 mM $ZnCl_2$
	1% (v/v) NP40
	0.1 mM PMSF
	1x Complete EDTA free (Roche)
TEV cleavage buffer	50 mM TRIS HCl pH 8.0
	0.3 M NaCl
	0.5 mM EDTA
	0.52 mM $ZnCl_2$
	0.1% (v/v) NP40
	0.1 mM PMSF
	1 mM DTT
Calmodulin binding buffer	50 mM TRIS HCl pH 8.0
	0.3 M NaCl
	0.1% (v/v) NP40
	1 mM MgOAc
	2 mM $CaCl_2$
	1 mM imidazole
	10 mM 2-mercaptoethanol
EDTA elution buffer	50 mM TRIS HCl pH 8.0
	0.3 M NaCl
	0.1% (v/v) NP40
	5 mM EDTA

GST Pull-down assay:

Binding buffer	25 mM HEPES pH 7.5 12.5 mM $MgCl_2$ 20% (v/v) glycerol 0.1% (v/v) NP40
Wash buffer	20 mM TRIS HCl pH 8.0 100 mM NaCl 1 mM EDTA 0.5% NP40

Senescence-associated β-Galactosidase (SAβ-Gal) Staining

SAβ-Gal fixing solution	2% formaldehyde 0.2% glutaraldehyde 97.8% PBS-A
SAβ-Gal staining solution (fresh)	1 mg/ml X-Gal 40 mM citric acid/Na_3PO_4 solution 5 mM K ferrocyanide ($K_4[Fe(CN)_6]•3H_2O$) 5 mM K ferricyanide ($K_3[Fe(CN)_6]$) 150 mM NaCl 2 mM $MgCl_2$
Citric acid/Na_3PO_4 solution	36.85 mM citric acid solution 126.3 mM Na_3PO_4

2.2 DNA techniques

Plasmid vectors were obtained from the laboratory's stock or provided by other laboratories (as acknowledged). cDNAs were obtained from the Medical Research Council Geneservice.

2.2.1 Plasmid vectors

The following vectors were used in this work for the following applications;

pCR2/pCR4	TOPO TA cloning of new cDNAs
pcDNA6	cDNA expression
pRSET A	HIS-fusion protein bacterial expression
pGEX 6P1	GST-fusion protein bacterial expression
pBABE	retroviral cDNA expression
pMARX	retroviral cDNA expression
pRS	retroviral shRNA knockdown
pMLP	retroviral shRNA knockdown
pLKO.1	lentiviral shRNA knockdown

2.2.2 PCR

In order to attach the desired restriction sites to an existing cDNA for subsequent ligation into a different vector, PCR was performed using primers including the new restriction sites. PCR was also performed with *de novo* cDNA produced by using the Superscript III reverse transcription (RT) kit (Invitrogen). The Pfu Ultra DNA polymerase (Stratagene) was used for its high fidelity. The general reaction includes 1X Pfu Ultra reaction buffer, 50 ng of DNA template, 0.2 mM dNTP mix, 1 μM of each primer and 2.5 U of Pfu Ultra and water up to 50 μl total volume. The general PCR cycle programme was:

1. 95 °C for 2 min (denaturation)
2. 95 °C for 30 sec (cyclical denaturation)
3. (primers Tm – 4) °C for 30 sec (primer annealing)
4. 68 °C for 1 min/kb of template (extension)
5. cycle to step 2 for 29-34 times

Generally after the PCR, Taq DNA polymerase was added and incubated at 72°C for one hour to create a poly A tail on the PCR product. This was needed for ligation into either the pCR2 or pCR4 vectors (Invitrogen) depending on the restriction sites

required. By using these vectors the PCR product was amplified to generate larger amount of cut insert for ligation.

In addition, PCR was used for site-directed mutagenesis of cDNAs. For this an existing plasmid was amplified using primers containing the altered base(s) and the following PCR cycle programme was used:

1. 95 °C for 1 min (denaturation)
2. 95 °C for 50 sec (cyclical denaturation)
3. (primers Tm – 4)°C for 50 sec (primer annealing)
4. 68 °C for 2 min/kb of template (extension)
5. cycle to step 2 for 17 times

Following the PCR program, the reaction was treated with Dpn1, which only digests dam methylated DNA that has been isolated from *E. coli* and therefore specifically removes the original template plasmid. 5 µl of this reaction was then used to transform competent bacteria as described next.

2.2.3 Restriction digestion, ligation and transformation

Plasmid DNA was digested with the relevant restriction enzymes to isolate a DNA fragment to use either as a recipient vector or as an insert to subcloned into another vector. Typically, 5 µg of DNA was used to prepare a cut vector and 5-10 µg of DNA was used to prepare a cut insert. Digestion was typically carried out in a final volume of 50 µl containing 1X of the appropriate restriction buffer (New England Biolabs) and 1 µg/ml BSA (New England Biolabs) and using 20 U of each restriction enzyme. Digestion samples were incubated for 1 h-O/N at 37 °C. Preparations of cut vector were generally dephosphorylated with 10U calf intestine phophatase (CIP) to reduce cut vector self-ligation.

Digested DNA samples were resolved by electrophoresis in a 0.5 – 2% agarose gel containing 0.5 µg/ml ethidium bromide. After sufficient resolution was achieved, the DNA fragments were visualised using a UV transilluminator (UVP BioDoc-IT) and the relevant pieces of DNA were excised and recovered using the QiaQuick gel extraction kit (Qiagen), according to manufacturer's instructions. Concentrations of cut vector and insert were either estimated by comparison of ethidium bromide staining

internsity, or were more precisely determined by measuring the DNA concentration using a NanoDrop spectrometer (Thermo Scientific). Generally, ligations were performed using a 3-5:1 ratio of cut vector to cut insert, in a 10 µl reaction containing 1X T4 ligase buffer and 10U of T4 ligase. A control reaction, using water instead of insert, was also performed. Ligations were generally performed for 1h at room temperature (rt) or O/N at 16°C. Ligation of PCR products into either pCR2 or pCR4 were performed in a similar manner.

Following ligation, 5µl of the reaction (or the total reaction volume for TOPO ligation) was used to transform competent DH5α bacteria (provided by Marc Rodriguez-Niedenführ) or One Shot TOP10 (Invitrogen) chemically competent cells. Bacteria were kept on ice for 30 min, heat-shocked for 45 sec at 42 °C and put back on ice for 2 min. 100µl of SOC medium was added next and the bacteria were put in a shaking incubator at 37 °C for a minimum of 15 min. The bacteria were plated on petri dishes containing LB-agar with 100µg/ml ampicilin, and 20 µg/ml X-gal if applicable (for blue-white selection of pCR2/pCR4 clones). Similarly, re-transformation of existing plasmids was performed using 1 µl of DNA.

2.2.4 Plasmid DNA preparation

Clones from transformation were grown up as 5 ml cultures in LB containing 100µg/ml ampicillin at 37 °C O/N. The next day, an aliquot was taken for possible culture expansion, and the remainder of the culture was centrigured to recover the bacterial pellet and submitted for mini-scale plasmid preparation (miniprep) by the LRI Equipment Park. Alternatively, the miniprep was performed using the QiaPrep Spin Miniprep Kit (Qiagen), according to manufacturer's guidelines. These minipreps were then analysed by restriction enzyme digestion to ascertain whether they contained the correct insert. Positive minipreps were then checked by DNA sequencing, as described below. A culture of a positive clones was stored in a 1:1 volume of glycerol at -80 °C.

To produce a plasmid DNA stock, clones were generally grown in a 100 ml culture at 37 °C O/N for a maxi-scale preparation of plasmid DNA (maxiprep). The cultures were pelleted by centrifugation at 4000rpm at 4 °C. The pellet was then processed using the HiSpeed Plasmid Maxi Kit (Qiagen), according to the manufacturer's guidelines. The DNA was eluted in water and the concentration was measured by NanoDrop spectrometer. Plasmids were stored at 4 °C.

2.2.5 DNA sequencing

Sequencing of DNA was performed using the ABI PRISM Dye Terminator Cycle sequencing kit (Applied Biosystems). This constitutes a PCR reaction containing a low level of labelled ddNTPs, which terminate the extension by the DNA polymerase. This produces a series of PCR products of different length from which the position and type of labelled base can be derived (performed by the LRI Equipment Park). The PCR reaction consist of 8 µl big dye terminator mix 3.1 (BDT), 200-300ng of template plasmid and 3.2 pmol of a sequencing primer, made up to a final volume of 20 µl with water. Commonly used sequencing primers flanking the multiple cloning sites of the vectors used are listed in Table 2. The PCR cycling programme used was as follows:

1. 96 °C for 1 min (denaturation)
2. 96 °C for 10 sec (cyclical denaturation)
3. (primers Tm – 4)°C for 5 sec (primer annealing)
4. 60 °C for 4 min (extension)
5. cycle to step 2 for 24 times

PCR reactions were then purified using the DyeEx 2.0 spin kit (Qiagen), according to manufacturer's instructions and vacuum dried. Derived sequences were interpreted using the MacVector software (Accelrys) and the online BLAST program (NCBI).

Primer name	Sequence
pBABE F1	CGTCTCTCCCCCTTGAACC
pBABE R2	CTGCCTGCTGGGGAGCC
pRS F	TACATCGTGACTGGGAAGC
pRS R	TAAAGCGCATGCTCCAGACT
T7	AATACGACTCACTATAG
BGH	TAGAAGGCACAGTCGAGG
SP6	CATACGATTTAGGTGACACTATAG
M13F	GTTTTCCCAGTCACGAC
M13R	CAGGAAACAGCTATGC

Table 2. Commonly used sequencing primers

2.3 Cell culture

All mammalian cell types used were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% (v/v) foetal bovine serum (FBS), which together will be referred to as 'medium'. Cells were grown as monolayers in humified air that containing 5% CO_2 at 37°C. Before reaching confluency, cells were washed once in versene pre-warmed to 37°C and then incubated with a 1:3 trypsin (Gibco):versene solution until cells would detach from the plastic surface. Medium was then added to neutralise the trypsin and cells were seeded in new culture dishes, generally at in a split ration between 1:2 and 1:6 and the population doubling doublings were recorded.

2.3.1 Cell strains/lines

The cell strains and cell lines used in this work are:

- BOSC23 Ad5-transformed human embryonic kidney (HEK) 293 cell line expressing gag-pol and env genes for ecotropic retrovirus packaging (Pear et al., 1993)
- 293T Ad5/SV40 large T-transformed HEK293 (ATCC number CRL-1573)
- DU-145 Prostate carcinoma (ATCC number HTB-81)
- MEF mouse embryonic fibroblasts isolated from C57BL/6 embryos (performed by LRI Cell Services)

- FDF Human foetal dermal fibroblasts (Brookes et al., 2004)
- Hs68 Human neonatal foreskin fibroblast (ATCC number CRL-1635)
- Q34 Human adult skin fibroblasts (Brookes et al., 2004)
- Leiden Human diploid fibroblasts derived from a patient who is homozygous for a deletion in exon 2 of *INK4A* (Brookes et al., 2002)
- MEF Mouse embryonic fibroblasts, wildtype derived from C57BL/6 mice by LRI Cell services and $Mph2^{-/-}$ mutant obtained from H. Koseki (Isono et al., 2005a)

To allow for infection by ecotropic retroviruses, the various strains of primary human diploid fibroblasts (HDFs) were infected with an amphotrophic retrovirus encoding the mouse basic amino acid transporter which functions as the cell surface receptor for mouse ecotropic retroviruses. Due to the switch to direct amphotrophic viral infection of the primary HDFs over the course of this work, the need for the ecotropic receptor was alleviated.

2.3.2 Storage, recovery and harvesting of cells

For harvesting and storage, cells were released from the culture dish by trypsinisation as previously described and resuspended in PBS. Cells were centrifuged at 1500g for 5 min. For freezing the cells, the cell pellet was resuspended in freezing medium (50% (v/v) FBS, 10% (v/v) DMSO) and transferred into cryoviles (Nunc) and frozen gradually in a freezing container (Nalgene) at -80 °C. Following freezing, cells were transferred to liquid nitrogen containers. For harvesting cells, the pellet was washed for an additional time in 1ml PBS, transferred to an Eppendorf tube and centrifuged at 1500g for 5 min. The resulting pellet was either lysed directly or stored at -20°C or snapfrozen and stored at -80 °C to optimally preserve protein-protein interactions. To recover frozen cells, the relevant freezing vile was swiftly transferred from liquid nitrogen to a 37 °C. waterbath and incubated until just thawed and then transferred to a culture dish containing medium at 4 °C.

2.3.3 Transfection of human cell lines

Approximately $2x10^6$ cells were seeded per 10cm dish the day before transfection. The next day, 20 µg of plasmid DNA was mixed with 437.5 µl of water and 62.5 µl $CaCl_2$ was added. The DNA sample was further mixed by pipetting and then transferred to 500µl of 2x HBS in a drop-wise manner without any mixing. This DNA:HBS mixture was then incubated for 30 min at rt, and then added to the cells to be transfected in a drop-wise manner without mixing. The cells were then incubated at 37 °C for 14-17 hours. The next day the medium was replaced with 10ml fresh medium. Gene expression was assayed 24-48 hours later.

For generating stocks of infectious virus, BOSC23 (ecotropic) or 293T (amphotrophic) cells were transfected as above. For producing amphotrophic virus, 10 µg of plasmid DNA was combined with 8 µg of pCG Gag-Pol (retrovirus) or pCMV Δ8.2 (lentiviral), and 2 µg of pCG VSV-G and used to transfect 293T cells. 14-17 hours after transfection, the medium was changed to half the original volume to concentrate the virus being produced.

For stable transfection, the plasmid DNA was first linearised by cutting at a unique restriction site. After 48 hours, the cells were selected with the respective antibiotic drug. Following selection the cells were seeded at a $1:10^4$ dilution to allow for colonies to arise from individual clones. Selected clones were recovered by trypsinisation of the colony and transferred to a 96-well culture plate. Clones were then gradually expanded and screened for expression of the transgene. The best expressing clone was used for further expansion.

2.3.4 Viral transduction of human cells

For retroviral or lentiviral transduction of HDFs, the cells were split 1:4 the day before infection. To harvest virus, medium collected from the packaging cells was filtered through a 0.45 µm Millix-HV filter (Millipore). The filtered virus-containing medium was either added directly to the HDF cultures or frozen at -80 °C, resulting in a 50% loss of viral titer upon thawing. After harvesting virus for ecotropic infection, 5 ml of fresh medium was added to the BOSC23 cells and a second viral titer was harvested 24 hours later for a secondary infection. For amphotrophic virus, the medium from the

293T cells was only harvested once. For lentiviral infection the harvested virus was diluted 1:2 – 1:5.

After 32 to 48 hours, the medium was changed to fresh medium containing the appropriate antibiotic drug. Puromycin was used at 0.5 μg/ml, hygromycin B at 50μg/ml and blasticidin at 5 μg/ml. Selection efficiency was monitored using an uninfected control plate of cells.

2.4 Protein biochemistry

2.4.1 Preparation of total cell lysate

Harvested cell pellets were lysed by the addition of 5X v/v ice-cold lysis buffer. Lysis was allowed to proceed for 10 min on ice. Lysates were cleared of nuclear derbis by centrifugation for 10 min. at 14,000 rpm at 4°C and supernatant was collected on ice. Lysis was generally performed with NP40 buffer, but also RIPA buffer was used for assessing protein levels by direct immunoblotting.

The protein concentration of lysates was determined by the Bradford assay using Protein Assay Dye Reagent (Bio-Rad). This concentrated solution was diluted 1:5 in water and 1 ml of the final solution was pipetted in a plastic cuvette (Sarstedt). A duplicate or triplicate 2μl of each lysate sample to be assayed was added to individual cuvettes. Similarly, serial dilutions of BSA standards of known concentrations (20 – 0.125 mg/ml) were added to individual cuvettes. Cuvettes were then briefly mixed by flicking. Samples were incubated for 15 min. Consequently the absorbance was measured at λ=595nm using a Ultrospec-2000 spectrometer (Pharmacia Biotech). Values were compared to a standard curve obtained from the BSA dilution series.

2.4.2 SDS-PAGE

Polyacrylamide gels (8 x 10 x 0.075 cm) were poured using the Protean system (Bio-Rad). The running part of the gel was cast from the mix of 375 mM TRIS HCl pH 8.8, 0.1% (v/v) SDS, 0.1% (v/v) APS and 0.04% TEMED and 8-15% acrylamide depending on the molecular weight of the protein of interest. Following casting, the solution was covered with water to promote polymerisation of the forming gel. When the running gel

was fully polymerised, the water on top was decanted. Next the stacking part of the gel was poured containing 5% acrylamide in 375 mM TRIS HCl pH 6.8, 0.1% (v/v) SDS, 0.1 (v/v) APS and 0.1% TEMED and a comb was inserted to form wells for sample loading.

Protein samples were prepared by adding 5X Laemmli buffer to a 1X final concentration and samples were adjusted to an equivalent final volume by using 1X Laemli buffer. The samples were then boiled for 10 min. Molecular weight size markers (7-175kDa) (New England Biolabs) were used for monitoring fractionation, used for size estimation and assisting in gel orientation for immunoblotting. Electrophoresis was performed in 1X running buffer in Protean tanks (BioRad) at a constant current of 30 mA per gel until desired fractionation was achieved.

For separating the samples obtained from tandem affinity purification (TAP) the NuPAGE system (Invitrogen) was used for higher resolution and better compatibility with downstream proteomic analysis. Therefore, the relevant samples were prepared using NuPAGE LDS sample buffer (Invitrogen) including 0.01 M DTT. The samples were loaded onto a pre-cast 4-12% NuPAGE BIS-Tris gel (Invitrogen) and run in NuPAGE MOPS running buffer containing 0.1% NuPAGE antioxidant (Invitrogen). Gels were run at 200 V for 50 min.

2.4.3 Immunoblotting

After separation by SDS-PAGE, proteins were transferred to Protran nitrocellulose membrane (Whatman) using the Protean system (BioRad). The membrane was soaked in transfer, placed over the top of the gel and 'sandwiched' between 3MM filter paper (Whatmann). The closed cassette was then inserted in the Protean transfer module, so the proteins would transfer from the negative to the positive electrode. Proteins were transferred at a 300 V for 75 min. Following this, the membrane was incubated with blocking buffer for 1 hour.

The primary antibodies commonly used in this work are listed in Table 3. These antibodies will subsequently be referred to by their source/origin. This list does not include the antibodies generated over the course of this work, which will be described in Chapter 3. Blocked membranes were incubated with primary antibodies against the protein/epitope of interest. Antibodies were diluted in blocking buffer, typically at 1 μg/ml. Generally 4 ml of diluted antibody was used per full-size membrane and

incubated for 1h at rt or O/N at 4°C. Following incubation with the primary antibody, the membrane was washed 3 times for 10 min in washing buffer. Depending on the primary antibody used, the membrane was incubated for 1h a 1:2000 dilution in blocking buffer of HRP-conjugated anti-mouse of anti-rabbit IgG (GE Healthcare). Following this incubation, the membrane was washed again 3 times for 10 min in washing buffer and then incubated with a 1:1 mix of ECL reagents 1 and 2 (GE Healthcare) for 1 min to intiate the chemiluminesence of HRP. Membranes were wrapped in Saran film (Dow Chemical), transferred to a Hypercassette (GE Healthcare) and exposed to Hyperfilm MP (GE Healthcare) in the dark for various time periods to capture the chemiluminescent signal.

For reprobing with another antibody, membranes were washed in PBS-A to remove ECL signals before incubating with a new antibody. For removal of antibody, membranes were washed 3 x 10 min in washing buffer and incubated with stripping buffer for 30 min at 50 °C. After that, blots were washed 3 x 10 min in washing buffer before applying the next antibody. In order to store membranes, they were covered in a thin layer of blocking buffer and wrapped in Saran film and stored at -20°C.

Epitope/protein	Name	Species and Type	Origin/Source
Human P16^{INK4A}	JC8	Mouse monoclonal	Koh & Harlow, LRI
Mouse p16^{Ink4a}	M-156	Rabbit polyclonal	Santa Cruz
CDK4/Cdk4	DCS-35	Mouse monoclonal	Neomarkers
FLAG	M2 (F3165)	Mouse monoclonal	Sigma
FLAG affinity gel	M2 (A2220)	Mouse monoclonal	Sigma
HA (HA.11)	16B12	Mouse monoclonal	Covance
CBX7	ab21873	Rabbit polyclonal	Abcam
BMI1	ab14389	Mouse monoclonal	Abcam
CBX8	LAST+GALD	Rabbit polyclonal	Kristian Helin, BRIC
Histone H3	ab1791	Rabbit polyclonal	Abcam
Histone H3K27me3	07-449	Rabbit polyclonal	Millipore
CBP	07-482	Rabbit polyclonal	Stratagene
6xHIS	4D11	Mouse monoclonal	CRUK
MYC	9E10	Mouse monoclonal	CRUK

Table 3. List of commonly used primary antibodies detailing their species, type and origin/source

2.4.4 Immunoprecipitation

Immunoprecipitation (IP) was typically performed on 300 µl volumes of cleared NP40 lysate and 10% of this volume was kept as input sample. For precipitations, generally 5 µg of purified antibody or 20 µl of unpurified serum and 20µl Protein A or Protein G sepharose beads (Thermo Scientific) were used. The samples were incubated on a rotating wheel at 4 °C for 2 h or O/N. The beads were pelleted at 2,000 rpm for 5 min at 4 °C and washed 3 times for 10 min in 1 ml of ice-cold NP40 lysis buffer. Immune complexes were released from beads by boiling in 2X Laemmli buffer or eluted with 10 µg of the respective peptide in a 7X bead volume of lysis buffer. Samples were either stored at -20°C or directly separated by SDS-PAGE.

2.4.5 Chromatin Immunoprecipitation

Cells to be used for chromatin immunoprecipitation (ChIP) were treated with a cross-linking reagent before they were harvested. To this end, 37% (w/v) formaldehyde was

added to the medium for a final concentration of 1% (w/v). After incubation at 37 °C for 10 min, the cross-linking was then blocked by addition of 125 mM glycine and incubation at rt for 5 min. Cells were then washed twice with PBS-A, harvested with a cell scraper in PBS-A and centrifuged for 5 min at 1,500 rpm. Cell pellets were lysed in ice-cold ChIP lysis buffer and incubated on ice for 10 min and nuclear pellets were obtained by centrifugation at 14,000 rpm at 4 °C for 10min. The nuclear pellet was washed with 1ml ChIP buffer and resuspended in 300 µl of IP buffer per 10^6 cells. The chromosomal DNA was then sheared using a BioRuptor waterbath sonicator (Diagenode) at high setting. Samples were sonicated in pulses of 30 sec with 30 sec intervals over three 10 min periods, where the ice in the waterbath was replaced after each period. Following sonciation, the samples were cleared by centrifugation at 14,000 rpm for 10 min at 4°C. The shearing efficiency was checked by reversing the crosslinking and RNase treatment (see below) using 5% of the chromatin sample and analysing the DNA fragment sizes on a 1% agarose gel. The target DNA fragment size range was 0.5 – 1 kb.

The protein concentration of the resulting chromatin sample was measured by the Bradford assay, as described in section 2.4.1. Generally 0.1-0.5 mg of chromatin was used per IP, depending on the affinity of the antibody used and a 10% input sample was retained. The IP samples were adjusted to 300 µl and 5 µg of antibody was added. As a negative control either rabbit unspecific IgG (for rabbit polyclonal antibodies) or mouse anti-HA monoclonal (for mouse monoclonal antibodies) was used. Following addition of antibody, the samples were incubated on a rotating wheel at 4°C O/N. The next day, samples were cleared by centrifugation to remove unspecific complexes and the top 90% of the sample was transferred to a new Eppendorf tube containing 20µl Protein A/G beads (Thermo Scientific) pre-washed in ChIP buffer. Samples were incubated on a rotating wheel at 4°C for a minimum of 30 min. The beads were then centrifuged at 500g for 5 min at 4°C and washed 3 times with 1 ml of ChIP buffer for 10 min at 4°C.

The antibody/protein/DNA complexes were eluted from the beads by adding 50 µl of elution buffer at rt and shaking on a vortexer for 15 min at highest speed. This elution step was repeated once more. The input sample(s) were included at this stage and adjusted to 100µl with elution buffer. NaCl was added to a final concentration of 0.2 M and the samples were incubated at 65°C for 4h - O/N to reverse the crosslinking. Next, RNase A was added to a final concentration of 0.1 mg/ml and samples were incubated for 30 min at 37°C. Finally, the DNA was purified on a QiaQuick purification

column (Qiagen). The relative enrichment for specific DNA sequences was determined by quantitative PCR (qPCR), as described in section 2.5.2.

2.4.6 Tandem Affinity Purification

Tandem affinity purification was initially conducted on a relatively small scale (~4 x 10^7 cells), but was subsequently adjust to much larger cell numbers for sufficient yield for downstream analysis. The essential steps are the same and adjustment for small- or large scale are indicated as appropriate.

Cells were lysed in the TAP lysis buffer. After clearing by centrifugation, the lysate was incubated with 30 µl (small scale) or 300 µl (large scale) fast-flow IgG sepharose beads (Amersham) at 4°C O/N. The next day, beads were pelleted at 1,000 rpm and washed 3 times in lysis buffer and a final wash in TEV cleavage buffer for 10 min at 4°C. Next, the beads were resuspended in 300 µl TEV buffer and 25 U (small scale) or 100 U (large scale) of TEV protease was added and incubated on a rocker for 90 min at rt. The beads were then pelleted at 1,000 rpm and the supernatant was collected. Next, the beads were washed 3 times with 300 µl Calmodulin binding buffer for 10 min at 4°C and supernatants were combined. The sample was then centrifuged briefly at 1,000 rpm and transferred to a new tube to ensure no carry over of IgG sepharose beads and 30 µl (small scale) or 300 µl (large scale) of pre-washed calmodulin sepharose beads (Strategene) were added and incubated on a rotating wheel at 4 °C O/N.

The following day, beads were washed 3 times for 10 min each with 1 ml ice-cold Calmodulin binding buffer at 4°C. The bound proteins were eluted by two 10 min incubations with 500µl EDTA elution buffer at 4 °C. The combined eluate was centrifuged briefly at 1000 rpm at 4 °C and pipetted to another tube to ensure no carry over of beads. For protein precipitation, trichloroacetic acid (TCA) was added to a final concentration of 10% and briefly mixed by flicking the tube. The samples were then incubated on ice 1h – O/N at 4 °C. Next, precipitated proteins were pelleted by centrifugation at 18,000g for 30 min at 4 °C. The pellet was washed twice with 1 ml of ice-cold acetone and finally air-dried. The resulting pellet was resuspended in 1x NuPAGE LDS sample buffer (Invitrogen) and either stored at -20 °C or directly used for SDS-PAGE.

2.4.7 Protein gel staining

Proteins were fractionated by SDS-PAGE as described in 2.4.2. The gel was then fixed for 30 min in fixing solution and Coomassie staining solution as described in section 2.1. was applied and incubated on a rotating platform for ≥2 h at rt. Next, the gel was destained by several washes in fixing solution at rt. For protein samples from small scale TAP, SDS-PAGE was performed using the NuPAGE system as described in section 2.4.6. Following general fixing, the gel was stained using the SilverQuest kit (Invitrogen) according to the manufacturer's protocol. Due to the high sensitivity from silver ion reduction, this staining allows detection of down to 0.3 ng of protein, but it is not compatible with the downstream mass spectrometry analysis services provided by the LRI. For protein samples from large scale TAP, gels were fixed in a solution of 12.5% (w/v) TCA and 3.5% (w/v) 5-sulfosalicyclic acid (Sigma) for 30 min and stained in Colloidal Coomassie Blue (Sigma) O/N at rt. Alternatively for higher sensitivity, gels were fixed in a solution of 50% methanol and 7% acetic acid, stained in SYPRO Ruby solution (Invitrogen) O/N at rt, and washed once in 10% methanol, 7% acetic acid.

2.4.8 LC/MS/MS Mass spectrometry

The gels stained with either SYPRO Ruby or Colloidal Coomassie Blue were used to excise visible bands. Gel slices were added to an Eppendorf tube containing 50 µl of Milli-Q-filtered water. These samples were snap-frozen and transferred to the LRI Protein Analysis and Proteomics service. The protein-containing gel slices were processed for mass spectrometry using the Janus liquid handling system (PerkinElmer). Briefly, the excised protein gel pieces were placed in the wells of a 96-well microtitre plate and destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, reduced with 10 mM DTT, and alkylated with 55 mM iodoacetamide. After alkylation, the proteins were digested with 6 ng/µL Trypsin overnight at 37 °C. The resulting peptides were extracted in 1% v/v formic acid, 2% v/v acetonitrile. The digest was analysed by nano-scale capillary LC/MS/MS using a nanoAcquity UPLC (Waters) to deliver a flow of approximately 300 nL/min. A C18 Symmetry 5 µm, 180 µm x 20 mm µ-Precolumn (Waters), trapped the peptides prior to separation on a C18 BEH130 1.7 µm, 75 µm x 100 mm analytical UPLC column (Waters). Peptides were eluted with a gradient of acetonitrile. The analytical column outlet was directly interfaced with a

Triversa nanomate microfluidic chip for mass spectrometric analysis (Advion). Mass spectrometric information was obtained using an orthogonal acceleration quadrupole-Time of Flight mass spectrometer (SYNAPT HDMS, Waters). Data dependent analysis was carried out, where automatic MS/MS was acquired on the 8 most intense, multiply charged precursor ions in the m/z range 400–1500. MS/MS data were acquired over the m/z range 50–1995. LC/MS/MS data were then searched against a protein database (UniProt) using the Mascot search engine programme (Matrix Science).

2.4.9 Recombinant protein expression

2.4.9.1 GST-fusion proteins

E. coli (BL21 rec-A, pLysS) were transformed with the pGEX 6P1 vector containing the gene of interest. A set of 6-10 clones were picked and analysed for efficiency of induction following addition of 1 mM IPTG. Briefly, bacterial pellets were lysed, lysates fractionated by SDS PAGE and evaluated by Coomassie staining. The best-expressing clone was used to set up a large culture of 1–5 l and grown in an incubator at 37 °C. The culture was induced with 1 mM IPTG at an OD at 600nm of 0.8, at which the culture is in log phase of growth. Generally the culture was induced for 4h at 37 °C. For induction at 16°C, the culture was cooled down for 30 min before addition of IPTG and the culture was grown O/N. The bacteria were then harvested by centrifugation at 4000g for 10 min at 4°C.

The pellet was resuspended in 5 volumes of lysis buffer with 1 mg/ml lysozyme and incubated for 15 min at rt. The lysate was then transferred to a glass centrifuge tube (Corex) and sonicated at highest intensity on ice for ≥6 pulses of 45 sec with 15 sec intervals until the lysate was no longer viscous. Lysate was then cleared by centrifugation at 14,000 rpm for 10 min at 4 °C and the supernatant was then collected. A 50% slurry of glutathione sepharose 4B beads (GS4B) (GE Healthcare) was added at 1 ml/l of bacterial culture, pre-washed in PBS. The lysate and beads were incubated for 30 min at rt or 2 hour at 4°C, depending on protein stability. Following incubation, beads were washed 3 times in 10X bead volume of ice-cold PBS for 10 min and centrifugation at 1000 rpm for 5 min at 4°C. The fusion protein was then eluted 3 times with 1ml GST elution buffer per ml of beads, by incubation for 10 min at rt and the

three eluates were combined. The purified protein was then dialysed against 50 mM TRIS HCl pH 8.0 to remove glutathione. For use in antibody generation, the fusion proteins were dialysed against PBS.

Alternatively, the protein was recovered from the GS$B beads by proteolytic cleavege. To this end, the beads were washed 3 times in 10X bead volume of cleavage buffer. Next, the beads were resuspended in 960µl of cleavage buffer/ml of beads and 40 µl (80 U) per ml of beads of PreScission Protease (GE Healthcare) was added and incubated for 4h at 4°C. Following incubation, the cleaved protein was recovered by centrifugation at 1000 rpm for 5 min at 4°C, while the GST and protease were retained on the beads.

The purified proteins were analysed by SDS-PAGE in combination with a series of BSA standards of known concentration (0.5 – 20 mg/ml). This gel was then stained with Coomassie and the BSA standards were used to roughly determine the concentration of the bands of interest.

2.4.9.2 HIS_6-fusion proteins

The pRSET A vector containing the relevant cDNA was used to transform BL21 bacteria. Colony selection, culture induction and bacterial collection was performed as for GST purification. Lysis was performed using 8 ml HIS purification buffer with 1 mg/ml lysozyme per liter of bacterial culture. Lysis and sonication were performed as described above. 1ml of NiNTA agarose beads (Qiagen) were placed in a purification column and washed once with 6 volumes of filtered water and twice with 6 volumes of HIS purification buffer. The lysate was then transferred to the purification column and incubated with gentle shaking for 30 min at rt. The column was then washed 6 times with 8ml of wash buffer (HIS purification buffer with HIS imidazole buffer to a final concentration of 20 mM imidazole). Protein was eluted by adding elution buffer (HIS purification buffer with HIS imidazole buffer to a final concentration of 250 mM imidazole), snapping off the cap and collecting 1 ml fractions. Individual fractions were analysed by SDS-PAGE including the BSA standards as described above.

2.4.10 In vitro binding assay

1 µg of each cDNA to be evaluated was in vitro transcribed and translated in the presence of ^{35}S-methionine using the TnT kit (Promega) according to the manufacturer's instructions. 20 µg of GST-fusion protein or GST alone was added to 20 µl of 50% GS4B bead slurry and made up to a total of 300 µl with binding buffer. These samples were incubated on a rotating wheel O/N at 4°C. Protein-bound beads were then placed in binding buffer containing 150 mg/ml BSA, 5% (v/v) FBS and 1 mM DTT on a rotating wheel for 10 min at 4 °C. 8 µl of the *in vitro* translated (IVT) samples were then added to the blocked beads and incubated for 1 h at rt with rotation. Beads were washed 4 times in ice-cold washing buffer and once in ice-cold 50 mM TRIS HCl pH 7.4. Beads were then boiled in 2X Laemmli buffer and separated by SDS-PAGE alongside 0.8 µl (10%) IVT input samples. Gels were fixed as described in section 2.4.7 and incubated with Amplify (GE Healthcare) according to manufacturer's instructions. Gels were then vacuum dried and the labelled protein were detected by autoradiography at -80 °C.

2.5 RNA techniques

2.5.1 RNA isolation

Total cellular RNA was isolated using the RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. The RNA was eluted in 30 µl of RNase-free water. The concentration of RNA was determined by Nanodrop spectrometer. For storage, purified RNA was snap frozen and stored at -80 °C.

2.5.2 Reverse transcription quantitative PCR (RT-qPCR)

RT was performed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems), according to the manufacturer's instructions. Typically 0.25 – 1.0 µg RNA was used in a 50 µl reaction primed with oligo dT. The reaction was incubated for 1 h at 37 °C on a PCR block.

In order to specifically detect the mature mRNAs in the cDNA mixture, primers were designed to amplify across an exon-exon junction. Further criteria for primer design were: size = 20-23mers, Tm ≥ 60°C, GC content = 50-60%, amplicon = 150 bp, and not more than 2 consecutive repeats of same base. Primers commonly used throughout this work are listed in Table 4. The qPCR reactions were prepared using the EXPRESS qPCR Supermix with Premixed ROX (Invitrogen), which is based on SYBR green incorporation and includes the ROX passive reference dye. 2 µl of the RT reaction and 0.2 µM of both primers were added to the qPCR Supermix at a 1X final concentration of. The reaction mixture was generally prepared per primer used and pipetted into a 96-well optical plate. The plate was loaded on the 7500 FAST Real-Time PCR System (Applied Biosystems) and the following PCR cycle protocol was executed:

1. 95°C for 15 seconds (hotstart of the polymerase)
2. 95°C for 3 sec (denaturation)
3. 60°C for 30 sec (combined annealing and extension)
4. read fluorescence
5. cycle to step 2 for 39 times
6. melting curve

In order to check if the signal observed originated from the specific amplicon produced by the primers, rather than possible primer-dimers, the PCR reaction was followed by melting curve analysis. The obtained data were analysed using the corresponding SDS software (Applied Biosystems).

The also DNA samples recovered in ChIP were also quantified in a relative manner using qPCR. The procedure is largely the same as for cDNA, other than that 4 µl of purified DNA was used. The primers used for qPCR for ChIP samples are largely based on previous published primers (Bracken et al., 2007) and are listed in Table 4 below.

Target	Sense Sequence	Antisense sequence
P16^{INK4A}	CGGTCGGAGGCCGATCCAG	GCGCCGTGGAGCAGCAGCAGCT
P15$^{INK\$B}$	TAGTGGAGAAGGTGCGACAG	GCGCTGCCCATCATCATG
ARF	CCCTCGTGCTGATGCTACTG	ACCTGGTCTTCTAGGAAGCGG
HPH1	GCACTCTCTCCAACATCTCCT	ACACCTCCTCTACACTCCAAC
HPH2	CCACTTACCAAGGATACCAAGA	GAGATGGGTGACAAGGGTTC
HPH3	GGAGTACTTCCCCTACCTAACG	GACTGACAGGAAGATGACGATG
CBX7	CGGAAGGGTAAAGTCGAGTATC	TCGGACCTCTCTTCCTATACCC
GAPDH	TGATGACATCAAGAAGGTGGTGAAG	TCCTTGGAGGCCATGTGGGCCAT

Table 4. List of primers used for amplification of cDNA from mature mRNA

Name	Sense sequence	Antisense sequence
PS1	GGAACCTAGATCGCCGATGTAG	TGTTTTACGCGTGGAATGCAC
PS1b	GCCCAACTCCACCAGATAGC	GCTCTGCAGCGTCGTGAT
PS1c	TCCATTTTCGAAGCCACCT	GGCTGCTGATGAAACAGCTAA
PS1d	CCTAAGGGATAAAGCTAGTGTCATAGG	TAGGCAGGGATAGGGGAGAA
PS2	GTGGGTCCCAGTCTGCAGTTA	CCTTTGGCACCAGAGGTGAG
PS3	GGAGCGATGTGATCCGTTATC	TGAAATCCCAATCGTCTTCCAC
PS3a	CTTCAAAAACAATTAGGAGGAGCTT	TGAAACATTGAAAACATAAGGATATTG
PS3b	ATTCAAAAATGAGGGGACTGAA	GAAATTTGGAACTCAAAGACACG
PS4	GCACTTGCCCTTCCAGGTATA	TGATAGTTCAAGGCCCTATGCC
PS5	CTCAAAGCGGATAATTCAAGAGC	AAGCCTTAAGAACAGTGCCACAC
PS6	ACCCCGATTCAATTTGGCAG	AAAAAGAAATCCGCCCCCG
PS7	AGAGGGTCTGCAGCGG	TCGAAGCGCTACCTGATTCC
PS8	GCCAAGGAAGAGGAATGAGGAG	CCTTCAGATCTTCTCAGCATTCG
PS9	CAAGCTTCCTTTCCGTCATGC	GCCAGAGAGAACAGAATGGTCAGAGCCA
PS10	ATCGCTGAGCGATGAAGGTAG	ATCACAAAAAGGAAAGGCAAG
PS11	TAGGAGGCCCCATTAAGCATAC	TGTAGTTGCCAGGAGTTGGAGG
PS12	TGTCTACCCAACACTTCCCTGC	AAGGCAAAGGTAACTGAAACGC

Table 5. List of primers used for amplification of the *INK4/ARF* locus.

Target	Sense Sequence	Antisense sequence
HPH1 exon 2	TATGAACGACAAGCAGTGCAG	GTCTGCTGCTGTGTAGTGCTG
HPH2 exon 5	CACGCTCTTCCCCAGCCCTGACT	CTGTCAGCTATGCCAGGCTT
HPH2c end	CAGCCGTTGCTCAGATAACTCA	CTGTATATAGAAGTCTCATGGCCTTT
HPH3 exon 1	GAAGCGGAATTTAAGGACCAT	CTGTACAGCATGTCGGTCTGA

Table 6. List of primers used for differential PCR of HPH transcripts

2.5.3 Absolute quantification using qPCR

The absolute expression levels of specific transcripts were determined by employing corresponding plasmid cDNAs of known copy number. A 10-fold dilution series of the plasmid DNA was made, ranging from 30 - 30,000 DNA molecules. Molecule number/µl was calculated using the formula;

$(9.1 \times 10^{11}$/ length of plasmid (kb)) x (plasmid concentration (ng/ml) / 1×10^{-6}) =
no of plasmid DNA molecules per µl.

The plasmid dilution series was piptetted in duplicate alongside the cDNA samples. The resulting C_T values were used to make a standard curve, from which a formula was derived in Excel (Microsoft) to calculate the number of specific cDNA molecules present in the cDNA samples

2.5.4 Sense/Antisense-specific RT-qPCR

This procedure largely follows the same steps as in the previous section, but using a sequence specific primer rather than oligo dT. By selecting sequence specific primers to drive reverse transcription it is possible to dictate whether the reaction is is measuring sense or antisense transcription in respect to the gene of interest. For subsequent qPCR, both the primer used in the RT reaction and the matching reverse primer were used (as listed together in Table 5).

2.5.5 RNA ChIP

In order to evaluate RNA-protein interactions within the nucleus, RNA-ChIP was employed. The fixing and harvesting of cells is the same as for ChIP described in section 2.4.5. The cell pellet was lysed in 5 volumes of ice-cold buffer A with NP40 and placed on ice for 10 min. Crude nuclei were pelleted by centrifugation at 5000rpm for 5 min at 4°C. The pellet was once washed in buffer A without NP40 and resuspended in buffer B and placed on ice for 10 min. Next, lysates were sonicated and cleared as described in section 2.4.5. The lysate was diluted in IP-buffer to a final volume of 1 ml per IP. A 10% aliquot was taken as input sample. 5µg of antibody and appropriate unspecific antibody control(s) (as described in section 2.4.5) were added and the samples were incubated on a rotating wheel at 4°C O/N.

The samples were transferred to washed 50 µl protein A/G agarose beads for collection of immune complexes by incubation on a rotating wheel for 2h at 4°C. Beads were pelleted by centrifugation at 1000 rpm for 2 min at 4°C. Beads were washed for 5 min each, once in low-salt wash-, high-salt wash- and LiCl wash buffers and twice in TE pH 8.0. Immune complexes were then eluted twice by addition of 250µl elution buffer (freshly prepared) and incubation in a rotating wheel for 15 min at 4°C. Eluates were combined and NaCl was added to a final concentration of 200mM and placed at 65°C for a minimum of 2h. Next, 20 µl of 1M TRIS HCl, 10µl 0.5M EDTA and 20µg proteinase K were added to each sample and incubated for 45 min at 42°C. Samples were then treated with TURBO DNase 1 (Ambion) for 30 min at 37°C. The RNA was recovered using RNeasy spin columns (Qiagen) were then used according to the manufacturer's RNA cleanup protocol, including an additional on-column DNase treatment with the RNase-free DNase set (Qiagen). RNA was eluted in 30 µl of RNase-free water and directly used for antisense specific RT-qPCR as described in section 2.5.4.

2.5.6 RNA interference

siRNAs targeting the gene of interest were either obtained from the Netherlands Cancer Institute (NKI) shRNA library or designed using the siRNA Target Finder (Genescript) online tool. Additional sequences were copied from the Sigma Mission collection (evaluated in pLKO only). The relevant sequences were then converted into the

following short hairpin template for direct cloning into the pre-cut retroviral vector pRetroSuper;

sequence reverse & complement sequence

F: 5' GATCCCC(…..n19.....)TTCAAGAGA(…..n19....)TTTTTGGAAA 3'

R: 5' AGCTTTTCCAAAAA(.....n19.....)TCTCTTGAA(.....n19....)GGG 3'

BglII HindIII

Or for cloning into the cut lentiviral vector pLKO.1;

sequence reverse & complement sequence

F: 5' 5' CCGG(……n19……)CTCGAG(……n19……)TTTTTTG 3'

R: 5' 5' AATTCAAAAAA(……n19……)CTCGAG(……n19……) 3'

Age1 EcoR1

The oligonucleotides were annealed by incubating 3 µl of each oligo at 1 mg/ml in 44 µl annealing buffer at 95 °C for 4 min and at 70 °C for 10 min. The samples were then allowed to cool slowly to allow annealing. Annealed oligos were then ligated into cut vector and used to transform competent bacteria, as described in 2.2.3. Positive clones were identified by restriction enzyme digestion and sequencing, as described in section 2.2.3 and 2.2.5. Cloned shRNAs were validated for knockdown of target sequence by qPCR or immunoblotting. The validated siRNA sequences used in this work are listed in Table 7.

Name	Target sequence	Position in cDNA	Vector
HPH1-1	CGACAAGCAGTGCAGGCTC	CDS nt100-118	pRS/pLKO
HPH1-2	GAAGCCAACTATGCTCGCG	CDS nt2539-2557	pRS
HPH1-3	GAGGAGGTGTACGAGTTTATT	CDS nt2827-2847	pLKO
HPH2-1	GGAATCTCAAGAAGAAGTA	CDS nt1784-1802	pRS
HPH2-2	CCAGAATGGTGAGAATAAA	CDS nt1656-1674	pRS/pLKO
HPH2-3	CTCAAGCTATGAGGAACCCTT	CDS nt2229-2249	pLKO
HPH3-1	CCAAGTCAGTCTCCTACTA	CDS nt1252-1270	pRS/pLKO
HPH3-2	AGCCAGAGCTACAGAATAA	CDS nt2279-2297	pRS/pLKO
CBX7-1[2]	GGATGGCCCCCAAAGTAC	CDS nt100-117	pRS
CBX7-2	GTATAGGAAGAGAGGTCCGAA	CDS nt204-224	pLKO

[2] Kindly provide by Jesus Gil, Imperial College, London

Table 7 siRNA sequences that were cloned into a shRNA format in either a retroviral or lentiviral vector or both.

2.6 Cell Biology Techniques

2.6.1 Senescence-associated β-Galactosidase Assay

Cells were plated at equal density the day before the assay was performed. The next day, cells were washed twice in PBS and then fixed for 3-5 min at rt with the SA-β-Gal fixing solution. Cells were washed in PBS twice and staining solution was added at 2 ml/35 mm dish. Dishes were incubated at 37 °C O/N. Resulting staining was captured by light microscopy.

3 The Generation of Antibodies against HPH2 and CBX7

3.1 Introduction

At the onset of this study there were no antibodies available to HPH2 and CBX7 from a commercial source. Therefore, an initiative was taken to produce monoclonal antibodies to both these proteins. Initially, a recombinant fragment of CBX7 was supplied to the internal CR-UK Monoclonal Antibody Service, but due to disappointing results the HuCAL recombinant antibody technology offered by AbD Serotec (Kidlington, UK) was employed to successfully create antibodies against both proteins. In addition, a number of polyclonal antibodies were raised in rabbits against CBX7 and all members of the HPH family. This chapter outlines the strategies used to generate and evaluate reagents that were deployed in the experiments described in later chapters.

3.2 Generation of CBX7 antibodies

3.2.1 Mouse monoclonal antibodies to a CBX7 fragment

Dating from the original demonstration that CBX7 can act as a repressor of *INK4A,* several members of the laboratory had tried to express recombinant CBX7 in *E. coli.* While it was possible to generate and purify a GST-tagged version of mouse Cbx7, analogous approaches with human CBX7 were consistently unsuccessful. As discussed later, a commercial laboratory (AbD Serotec) also failed to produce recombinant CBX7 in their optimised bacterial expression system. To try to surpass the problems with recovery of the full-length protein, the human CBX7 cDNA was divided into five separate fragments based on the functional domains of the proteins (Figure 3.1A) and each fragment was expressed as a fusion protein with an N-terminal GST-tag. As illustrated in Figure 3.1B, only three of the fragments were expressed at workable levels. Due to the high immunogenicity of GST, an additional obstacle was the need to remove this tag prior to immunisation of mice. Following cleavage of the linker sequence with Precision Protease, only fragment 1 was recovered in sufficient quantities

to proceed (Figure 3.1C). Although recognising that this fragment was not ideal as it contains part of the chromodomain that is highly conserved in other CBX proteins, the antigen was provided to the CR-UK Monoclonal Antibody Service for mouse immunisation (see Figure 3.2A).

After the B-cell fusion stage, 61 hybridomas were found to be positive for the CBX7 antigen, as measured by ELISA. The supernatants of the hybridomas were then tested by immunoblotting, to confirm reactivity with fragment 1. 20 out of the 61 hybridomas tested were positive in this assay and based on the strength of signal, 8 were selected for single cell cloning and expansion. Unfortunately, upon screening of single cell clones of the positively identified hybridomas, the only consistently observed reactivity was to a 73 kDa band in the antigen sample, presumably corresponding to a contaminant from the purification. In an attempt to overcome this problem, a HIS_6-tagged version of CBX7 fragment 1 was produced in the hope that the different purification method would eliminate this particular contaminent (see B). Frozen clones from positive hybridomas were recovered and the supernatants were re-screened with HIS_6-CBX7 fragment 1 by ELISA, but no significant reactivity was found. It was therefore concluded that the positive clones had been lost at the stage of single cell cloning, perhaps by picking too few clones, or that antibody production had switched off in the frozen cells. Despite the substantial input of effort, it was decided not to further pursue this approach.

3.2.2 Human recombinant antibodies to CBX7 fragments

Instead of using the CR-UK monoclonal service, it was decided to commission recombinant antibodies from AbD Serotec (Kidlington, UK). The Human Combinatory Antibody Library (HuCAL) technology is based on screening an antigen against a library of over 1.5×10^{10} antibody specificities in Fab format displayed by phage. This allows the identification of candidates that are then further validated by ELISA. Positive clones can then be used to generate recombinant human antibodies in a variety of different formats. The standard format is a bivalent antibody with an artificial dimerization site and carrying the MYC and HIS_6 epitopes to facilitate detection. The screen could be designed to offer differential ELISA screening with the GST fusion protein and GST alone. So, as there was no need to remove the GST-tag, both CBX7 fragments 1 and 4 (Figure 3.2C+D) were available in sufficient amounts to be screened.

Regrettably, AdB Serotec made a pro-active decision to perform the initial panning with fragment 1 and then to check the positive clones for reactivity against fragment 4, whereas the reverse strategy would have reduced the aforementioned problem of chromodomain conservation in CBX7 fragment 1. In fact, CBX7 fragment 4 would have offered 58 unique residues of a lower degree of conservation in comparison to the 47 unique residues of CBX7 fragment 1. As shown in Figure 3.3A, eleven clones were obtained in the screen, of which four were also positive against fragment 4, implying that they recognise the shared region of amino acids 48 to 81 of full-length CBX7.

All eleven of the recombinant antibodies were able to IP FLAG-tagged human CBX7 in transiently transfected 293T cells (see Figure 3.3B). It is however not yet clear whether they will support chromatin immunoprecipitation (ChIP). A complicating issue for this application is the use of MYC and HIS_6 epitopes in the antibody backbone. As endogenous c-MYC is present at the promoters of many genes (Fernandez et al., 2003), secondary precipitation with the 9E10 monoclonal antibody could result in a high background. Similarly, many endogenous proteins are known to bind to Ni-NTA beads. As illustrated in Figure 3.3C, only one of the antibodies tested by ChIP showed significant enrichment over beads alone with the anticipated discrimination between the *INK4A* and *ARF* promoters. This has yet to be confirmed in follow up experiments. Enthusiasm for pursuing these antibodies was diminished by the clear evidence for cross reactivity with other members of the human CBX family (Figure 3.3D, performed by Marc Rodriguez-Niedenführ). This is likely to be caused by conservation of the epitope, assumed to be located in the chromodomain. However, while none of the antibodies can be regarded as monospecific, several show a strong preference for CBX7. Moreover, given the size differences between members of the CBX family, it should be possible to use the more broadly reactive clones as pan-CBX reagents in immunoblotting.

3.2.3 Rabbit polyclonal antibodies to CBX7 peptides

Prior to the start of this project, there had been a number of attempts in the laboratory to generate rabbit antibodies against synthetic peptides, based on the primary sequence of human CBX7. None of these reagents proved satisfactory. There was however a commercial source (Abcam) for a polyclonal antibody that was found to work well in immunoblotting, IP and ChIP. However, as commercial reagents can suffer from batch-

to-batch variation and significant quantities of such an antibody were desired, one further attempt was made to obtain a ChIP grade antibody specific for CBX7. Following the report of a ChIP grade CBX7 antibody by the Helin lab (Bracken et al., 2007), further attempts were made to use the so-called RELF peptide sequence (Table 8) for immunization. Evaluation of the final bleeds from the two rabbits immunised (designated #3 and #4) are shown in Figure 3.4. Although both antisera were able to detect FLAG-tagged CBX7 when over-expressed in 293T cells, (Figure 3.4A), they did not work convincingly for immunoprecipitation of the same protein (Figure 3.4B). Similarly, the C-RELF4 antiserum did not support ChIP under conditions in which a commercial CBX7 antibody (Abcam ab21873) showed significant enrichment at the *INK4A* locus using primer sets described in more detail in Chapter 5 (Figure 3.4C).

3.3 Generation of HPH2 Antibodies

3.3.1 Human recombinant antibodies to full length HPH2b

In anticipation of making monoclonal antibodies against HPH2, plasmid vectors were generated encoding the short isoform of HPH2, HPH2b (discussed in more detail in Chapter 4), as a fusion protein to either GST or HIS_6. Upon testing these constructs it was found that the achievable expression was low. In order to produce sufficient quantities of recombinant protein, the purification was scaled up to 5 liters and the induction was at 16º C overnight in order to allow for stable expression of protein. The purified protein consistently showed to have degradation products, as confirmed by immunoblotting for HPH2 (using the 1615 antiserum described below) (see Figure 3.5A). It was decided to use the recombinant protein for a HuCAL phage display screen. However, the methodology of AbD Serotec required the absence of primary amines. This meant that the existing preparations of GST-HPH2b had to be dialysed to remove TRIS from the buffer, but this simple step resulted in almost total loss of the protein. To circumvent this problem, focus was shifted to HIS_6-HPH2b, as the buffers used in the purification procedure do not contain primary amines. The final sample contained a rather low concentration of fractions of the protein as detectable by Coomassie staining, but the full-length protein could be detected in the sample by immunoblotting for HIS_6 and HPH2 (see Figure 3.5B).

The HuCAL screen yielded eight positive clones (Figure 3.6A), and the derived bivalent antibodies were subsequently tested for their ability to detect HIS_6-HPH2b by immunoblotting. Most antibodies recognised the antigen to a varying degree but the best signal was obtained for antibody #74 (Figure 3.6B), which matches to the affinities measured by ELISA by AbD Serotec. When tested for their ability to immunoprecipitate HA-tagged HPH2 from 293T cells, it was found that the procedure produced a very high background (Figure 3.6C). This was not resolved by pre-clearing with beads. This could possibly be due to the aforementioned limitations of the epitopes on these antibodies. Further work will be required to evaluate the full potential of these antibodies and it might be possible to reduce the background in IP by re-engineering the clones with different epitope tags, such as HA or FLAG epitope, which have not demonstrated any background for the IP of HPH2.

3.3.2 Rabbit polyclonal antibodies to HPH2 peptides

Because of the heavy demand on peptide synthesis and antibody production that ensued from the characterization of PRC1 complexes, the laboratory had taken the decision to order custom made peptide antibodies against HPH2 from Eurogentech (Seraing, Belgium). The strategy was to inoculate two rabbits with a combination of two synthetic peptides (see Table 8) and to affinity purify an aliquot of serum on the cognate peptides (E.85 and E.85) or on protein A sepharose. The evaluation of the resulting antibodies (Figure 3.7A) indicated that the 1615 antiserum was the more effective, as judged by immunoprecipitation of HA-tagged HPH2 overexpressed in 293T cells, but that either peptide-purified (E.84 and E.85) or IgG-purified serum (IS) offered little improvement compared to the crude serum (S). As will be discussed in chapter 5, the 1615 antisera were evaluated for the ability to confirm direct binding of HPH2 to the *INK4A* promoter region by chromatin IP (ChIP), showing a 3-fold enrichment compared to the non-specific rabbit IgG control (Figure 3.7).

Both of the antigenic peptides were from the amino terminal region of HPH2 that is common to the long and short isoforms of the protein (discussed in more detail in Chapter 4). Using the 1615 antisera an endogenous band of approximately 48 kDa could be detected by direct immunoblotting of untransfected 293T cells (Figure 3.7C). However, in parallel immunoblots of three strains of primary human fibroblasts (FDF, Hs68 and Q34), there was no equivalent band. Instead, the 1615 antibody detected

bands of 60, 55 and 36 kDa, The predicted molecular weight of HPH2b, based on its primary amino acid sequence is 35,760Da (as quoted in the Uniprot database), which corresponds precisely to the lower band. However, our favoured interpretation at this stage is that the three bands detected in primary fibroblasts correspond to degradation products from the long isoform, HPH2a. This is based on the appearance of similar bands (aligned by blue dotted lines) in 293T cells transiently transfected with an HA-tagged version of HPH2a (Figure 3.7D). Interestingly, the 48kDa band observed in untransfected 293T matches to that observed for ectopically express HA-tagged HPH2b (Figure 3.7D, as aligned by a red dotted line). Similar results were obtained using the 1616 antibody (data not shown). Nevertheless, these experiments need to be repeated, preferably including a sample in which HPH2 is knocked down, before drawing a conclusion.

3.4 Polyclonal antibodies against other HPH proteins

As the project developed, it became increasingly obvious that the other members of the HPH family, HPH1 and HPH3, were implicated in the regulation of *INK4A*, exacerbating the need to have reagents for these proteins as well. To this end, the LRI Peptide Synthesis Service was asked to design and synthesize appropriate peptides (see Table 8) and rabbits were inoculated with mixtures of these peptides. At the time of writing, the resulting antibodies are being evaluated by other members of the laboratory (Marc Rodriguez Niedenführ and Julie Stock) and so far the antibodies against HPH1 and HPH3 are looking promising for use in immunoprecipitation, whereas the anibody against HPH3 also shows initial efficacy in the ChIP protocol.

3.5 Discussion

The original aim was to obtain monoclonal antibodies against CBX7 and HPH2. Unfortunately, the project from the LRI's mouse monoclonal service was unsuccessful. Hence the projects were transferred to AbD Serotec (Germany). Although a number of high affinity antibodies were derived against CBX7, these were not found to be mono-specific due to the high conservation of sequence of the antigen used, with respect to other PC family members. Therefore these antibodies only have limited application.

Furthermore the antibodies against HPH2 produced overall weaker signals when used for immunoblotting and so far have faced technical barriers for use in immunoprecipitation-based techniques. The polyclonal antibodies produced against HPH2, and particularly 1615, were successful in the evaluation for immunoprecipitation.

The lack of reliable antibodies against the human CBX and the HPH proteins, whether from commercial suppliers or other academic groups, and the difficulties encountered in generating reagents in-house, represented serious barriers to progress in the early phases of this work. Although there has been no systematic study, most anecdotal reports imply that PcG proteins are difficult to express as single recombinant proteins. Thus, despite taking advice from structural biologists accustomed to optimising expression in *E. coli*, the laboratory still faces a challenge in characterizing the human homologue of CBX7.

Mouse Cbx7 presents significantly increased stability over its human counterpart and can be employed to study PcG regulation of human *INK4A* (Gil et al., 2004; Maertens et al., 2009). We have noted that human CBX7 contains a region of 93 residues that is not present in the mouse homologue (Figure 3.1A). An MRes student in the laboratory has subsequently confirmed that these residues are in part responsible for the reduced stability of human CBX7 compared to mouse Cbx7. However, use of mouse Cbx7 is not ideal, as possible interactions with the missing 92 residues may play an important role in regulation of human CBX7 in the human system (Nicholls, 2006). It is therefore needed to further investigate the underlying cause of instability of human CBX7 and to optimise its expression, possibly by the modification of certain residues in the unique 93 amino acid domain. The resulting protein could then be used in future efforts to develop potent monoclonal antibodies against CBX7.

As an alternative to bacterial expression, preliminary efforts have been undertaken by the laboratory to turn to baculovirus expression systems. Baculovirus has been successfully used by other research groups to reconstitute active complexes of PRC1 proteins (Elderkin et al., 2007). In our laboratory, HPH2b and CBX7 were recently cloned into the BaculoDirect Baculovirus Expression System (Invitrogen) using Gateway Technology (performed by Tomas Racek). However, as baculovirus expression is not currently in routine use by the laboratory, no further progress has been made. It therefore remains to be evaluated whether the use of baculovirus would enhance expression of these two proteins.

The phage display approach has the potential to produce virtually unlimited quantities of recombinant antibody, rapidly and without problem of clonal instability. In principle, these engineered antibodies can have the same affinities as authentic antibodies, but the limited experience in the laboratory with the HuCAL antibodies suggests that epitope tags have problems of producing high background. Alternatively to changing the epitopes, it is possible to circumvent these problems by reconstruction of the full antibody rather than the artificial bivalent format, but this option has not yet been evaluated by our laboratory. These technical issues need to be further assessed, especially for the development of ChIP-grade antibodies. The amounts of antibodies used in large-scale or sequential ChIP experiments currently conducted by the laboratory would benefit from a more economical and stable source of antibodies.

In contrast, peptide immunisation of rabbits has succeeded in generating ChIP grade antibodies for HPH2 and HPH3, which open up exciting possibilities for future research, such as to dissect the co-localisation on chromatin. To much disappointment, the validated peptide from CBX7 did not lead to the generation of potent antibodies. It should be considered whether any difference in conjugation method, immunisation protocol or rabbit species exists, which may account for the difference in antibodies generated.

As time has passed since the start of this project, the range of commercial antibodies available has expanded but it is still far short of supporting a comprehensive study of human PRC1 complexes. Reagent generation and evaluation will therefore be a continuing challenge to the field.

Target	Name	Sequence	Designed by	Rabbit
HPH2	EP053084	CLPOKTKDTKKQPTGT	Eurogentech (Seraing, Belgium)	1615+1616
HPH2	EP053085	STSRRRQGQRDLELC	Eurogentech (Seraing, Belgium)	1615+1616
CBX7	C-RELF	KKEPPRGPNLESHSHRREL FLQEPPC	Helin Group (BRIC, Denmark)	MRO-3/4
HPH2a	HPH2a 1-18	MENELPVPHTSSAC	O'Reilly Group (LRI)	MRO-5/6
HPH2a	HPH2a 450-467	HTSAVILQLQPASPPQQC	O'Reilly Group (LRI)	MRO-5/6
HPH2	HPH2a 740-750	CSDNSSYEEP	O'Reilly Group (LRI)	MRO-7/8
HPH2	HPH2 841-858	CKLGPALKIYARISMLKDS	O'Reilly Group (LRI)	MRO-7/8
HPH1	1-16C	METESEQNSNSTNGSSC	O'Reilly Group (LRI)	MRO-25/26
HPH1	304-326	CPRKGTGVVQPLPAAQTV TVSQG	O'Reilly Group (LRI)	MRO-25/26
HPH3	304-326C	SLIKHQQIPLHSPPSKVSHH QLIC	O'Reilly Group (LRI)	MRO-27/28
HPH3	742-755	INSVCVQPELQNNT	O'Reilly Group (LRI)	MRO-27/28

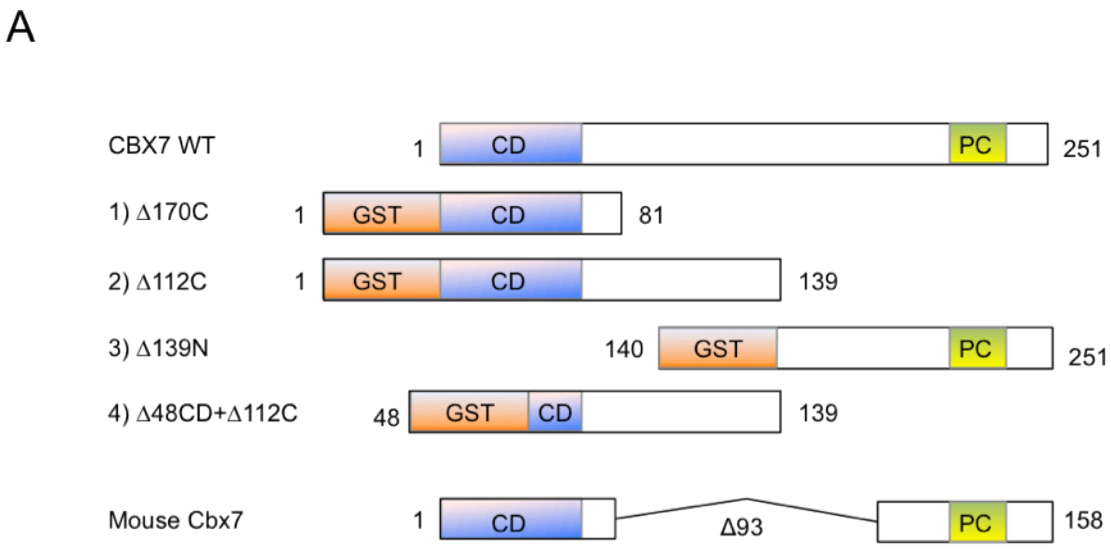
Table 8 List of peptides used for generation of rabbit polyclonal antibodies

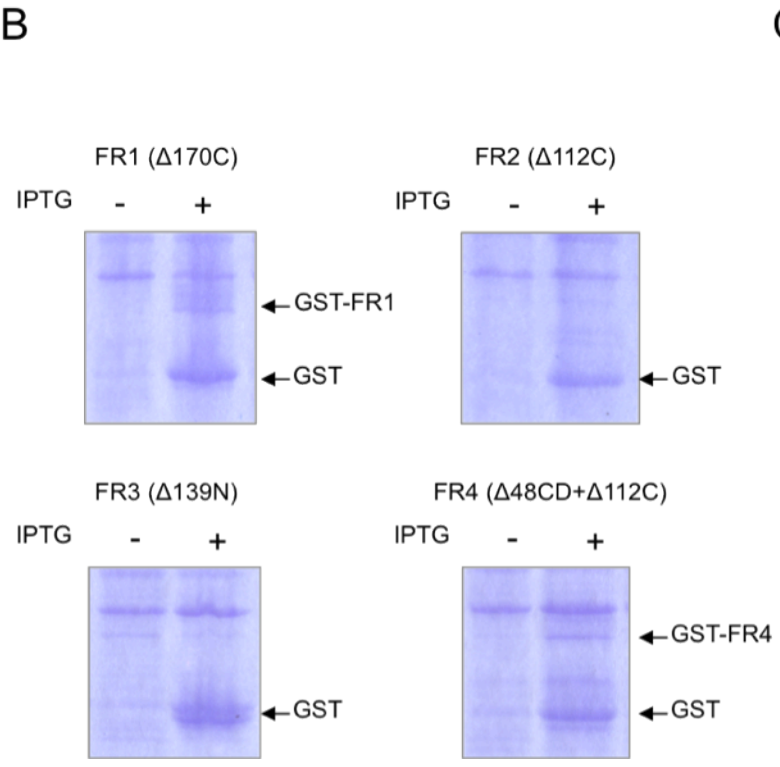
Figure 3.1 Expression of CBX7 Fragments

A) Overview of wildtype human CBX7, highlighting the location of the chromodomain (CD) and polycomb domain (PC), the 4 CBX7 fragments that were cloned to an N-terminal GST tag for purification, and the composition of mouse Cbx7 relative to the human form in terms of sequence conservation.

B) The recombinant GST-CBX7 fragments were tested for expression in *E. coli.* 20 ml cultures were induced for 4 h at 37 °C with 1 mM IPTG. The bacteria were lysed, lysates were fractionated by SDS-PAGE and resulting gels were stained with Coomassie staining solution. The presented induction for each CBX7 fragment corresponds to either the best (detectable expression) or a representative clone (undetectable expression) of a set of 5 clones per protein.

C) Evaluation of stability of CBX7 fragments following TEV cleavage. A 1 l culture of the clones for GST-CBX7 fragment 1, as displayed in B) was induced, harvested and used for GST purification and TEV cleavage as described in section 2.4.9.1. The resulting samples were fractionated by SDS-PAGE and the resulting gels were stained with Coomassie staining solution.


A
CBX7 WT
1
CD
PC
251
1) Δ170C
1
GST
CD
81
2) Δ112C
1
GST
CD
139
3) Δ139N
140
GST
PC
251
4) Δ48CD+Δ112C
48
GST
CD
139
Mouse Cbx7
1
CD
Δ93
PC
158


B
FR1 (Δ170C)
IPTG - +
GST-FR1
GST
FR2 (Δ112C)
IPTG - +
GST
FR3 (Δ139N)
IPTG - +
GST
FR4 (Δ48CD+Δ112C)
IPTG - +
GST-FR4
GST

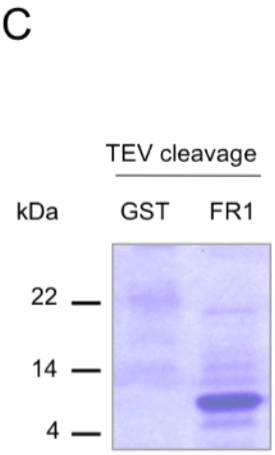

C
TEV cleavage
kDa
GST
FR1
22
14
4

Figure 3.2 Overview of the CBX7 antigens

A) CBX7 fragment 1 as derived from TEV cleavage following purification from 2 l of culture induced for 4 h at 37 °C as described in section 2.4.9.1. This protein was submitted as antigen to the CR-UK Monoclonal Production Service.

B) HIS_6-CBX7 fragment 1. In an attempt to resolve issues with cross-reactivity of antibodies to a contaminent in A, CBX7 fragment 1 was cloned with an N-terminal HIS_6 tag. A set of clones were evaluated for expression levels following 4 h induction at 37 °C. by 1mM IPTG. 2 l of the best clone were used for purification on NiNTA Superflow beads (Qiagen) and eluted using 3 M imidazole as detailed in section 2.4.9.2. This antigen was used as a secondary antigen to that described in A, and employed by the CR-UK Monoclonal Production Service to re-check the clones of positively identified hybridomas.

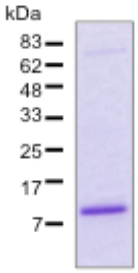
C) GST-CBX7 fragment 1. Due to higher stability in comparison to the cleaved fragment descibed in A, only 1 l culture was induced for 4 h at 37 °C and used for GST purification as described in section 2.4.9.1. This sample was supplied as antigen to AbD Serotec for screening human recombinant antibodies.

D) CBX7-CBX7 fragment 4. As the GST tag of CBX7 fragment 4 was found to greatly improve stability, this was supplied as an additional antigen to AbD SeroTec. 3 l of culture induced, purified and dialysed as described in C. This sample was supplied as antigen to AbD Serotec and was used for ELISA screening of human recombinant antibodies identified with the antigen in C.

A

CBX7 fragment 1

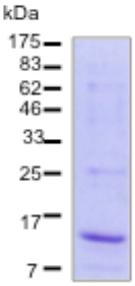
Molecular weight: ~9.7 kDa
Sequence:
ELSAIGEQVFAVESIRKKRVRKGK
VEYLVKWKGWPPKYSTWEPEEHI
LDPRLVMAYEEKEERDRASGYRK
RGPKPKRLLLQ
Concentration: ~1mg/ml
Quantity: ~1mg
Buffer: 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT



B

HIS_6-CBX7 fragment 1

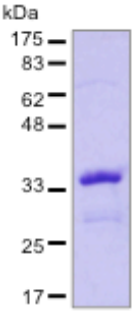
Molecular weight: ~35kDa
Sequence:
HIS_6+ELSAIGEQVFAVESIR
KKRVRKGKVEYLVKWKGW
PPKYSTWEPEEHILDPRLV
MAYEEKEERDRASGYRKR
GPKPKRLLLQ
Concentration: ~0.8mg/ml
Quantity: ~0.8mg
Buffer: 3M Imidazole pH 6.0



C

GST-CBX7 fragment 1

Molecular weight: ~35kDa
Sequence: GST+
ELSAIGEQVFAVESIRKKRV
RKGKVEYLVKWKGWPPKY
STWEPEEHILDPRLVMAYE
EKEERDRASGYRKRGPKP
KRLLLQ
Concentration: ~1.5mg/ml
Quantity: ~1.5mg
Buffer: PBS



D

GST-CBX7 fragment 4

Molecular weight: ~38kDa (full length)
Sequence: GST+
LDPRLVMAYEEKEERDRASGY
RKRGPKPKRLLLQRLYSMDLR
SSHKAKGKEKLCFSLTCPLGS
GSPEGVVKAGAPELVDKG
Concentration: ~1mg/ml (full-length)
Quantity: ~1mg
Buffer: PBS

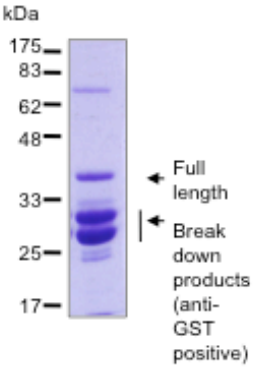


Figure 3.3 CBX7 Recombinant Human Antibodies Evaluation

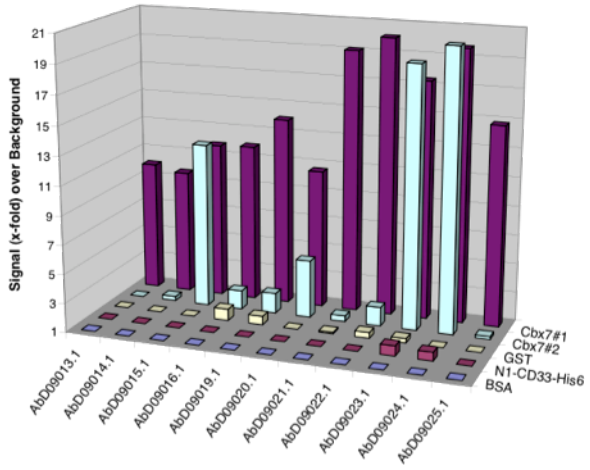
A) ELISA (as performed by AbD Serotec) was used to measure the relative affinity of the clones identified from the screen of CBX7 fragment 1 for recognition of both CBX7 fragments, as well as negative controls GST, N1-CD33-HIS_6 and BSA.

B) 293T cells transfected with pcDNA6 FLAG-CBX7 were lysed with NP40 buffer. The lysate was then incubated with 10 μg of all AbD Serotec CBX7 antibodies and precipitated using anti-Myc (9E10) and Protein A/G agarose. The immunoprecipitated samples and an input of 10% of the orginal lysate (INP) were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with anti-FLAG.

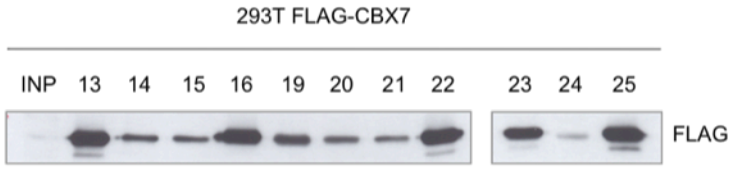
C) ChIP was performed as described in section 2.4.5 with the chromatin of mid-lifespan FDF cells (PD29) and the 5 antibodies of AbD Serotec CBX7 displaying the strongest IP capacity in B. An anti-Myc secondary antibody (9E10) was used to purify the immunecomplexes on Protein A/G-agarose. The purified DNA was analysed by qPCR using primers specific to the *INK4A* exon 1α (PS7) or to *ARF* exon 1β (PS2).

D) 293T cells overexpressing the different human PC proteins in pcDNA6 FLAG were lysed with NP40 buffer and used to IP for the FLAG-epitope to purify and concentrate the protein. IP samples were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with the AbD Serotec CBX7-13 antibody at a concentration of 5 mg/ml. Antt-human Fab-HRP was used as a secondary antibody in a 1:2000 dilution.

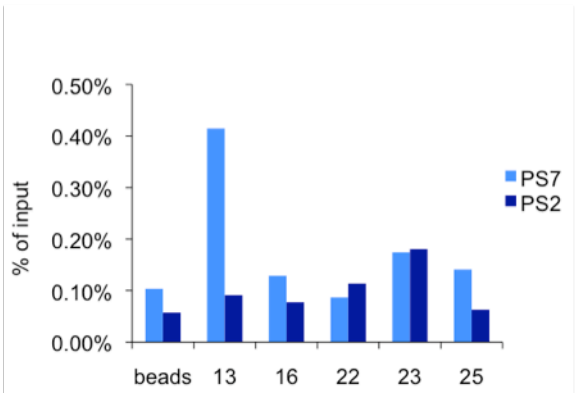
A


Signal (x-fold) over Background
21
19
17
15
13
11
9
7
5
3
1
AbD09013.1
AbD09014.1
AbD09015.1
AbD09016.1
AbD09019.1
AbD09020.1
AbD09021.1
AbD09022.1
AbD09023.1
AbD09024.1
AbD09025.1
Cbx7#1
Cbx7#2
GST
N1-CD33-His6
BSA

B


293T FLAG-CBX7
INP
13
14
15
16
19
20
21
22
23
24
25
FLAG

C


% of input
0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
beads
13
16
22
23
25
PS7
PS2

D

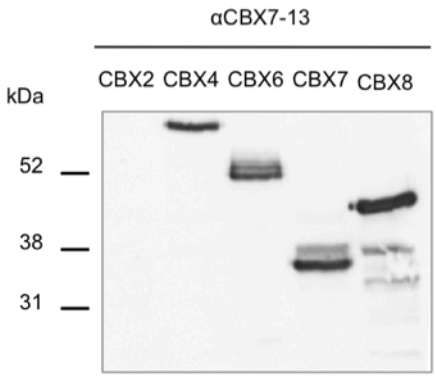

αCBX7-13
CBX2
CBX4
CBX6
CBX7
CBX8
kDa
52
38
31

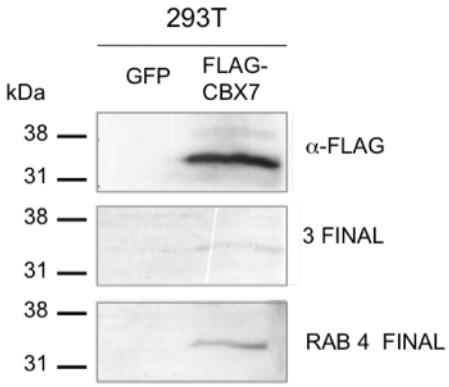
Figure 3.4. CBX7 Polyclonal Antibodies Evaluation

A) 293T cells transfected with pcDNA6 FLAG-CBX7 or pcDNA6 GFP were used to evaluate the antisera from rabbit 3 and 4, which were immunised with the C-RELF peptide. The cells were lysed in RIPA buffer. Protein was fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with the antisera from rabbit 3 or rabbit, 4 or with an anti-FLAG as control.

B) 293T cells transfected with pcDNA6 FLAG-CBX7 were lysed with NP40 buffer. The lysate was then incubated with 20µl of antiserum from rabbit 4. The resulting immunoprecipitated sample and an input of 10% of the orginal lysate used (INP) were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with anti-FLAG.

C) Chromatin Immunoprecipitation (ChIP) was performed as described in section 2.4.5 using chromatin from mid-life HS68 cells (PD45). The chromatin was incubated with 20µl of CRELF antiserum from rabbit 4 for with antibodies for CBX7 (Abcam) or unspecific rabbit IgG. The purified DNA was analysed by qPCR using primers specific to the *INK4A* promoter region (PS6 and PS7) or the *INK4B* exon 1 (PS1).

A


293T
GFP
FLAG-CBX7
kDa
38
31
α-FLAG
38
31
3 FINAL
38
31
RAB 4 FINAL

B


293T FLAG-CBX7
kDa
INP
CRELF4 IP
38
31
α-FLAG

C

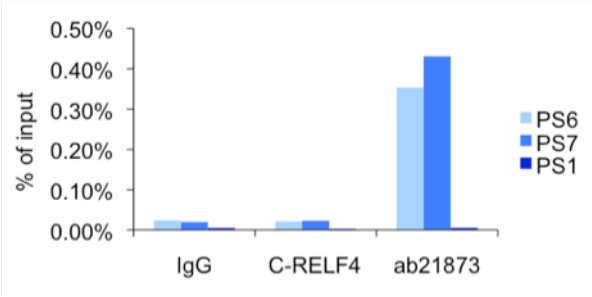

0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
% of input
PS6
PS7
PS1
IgG
C-RELF4
ab21873

Figure 3.5 Overview of HPH2 Antigens.

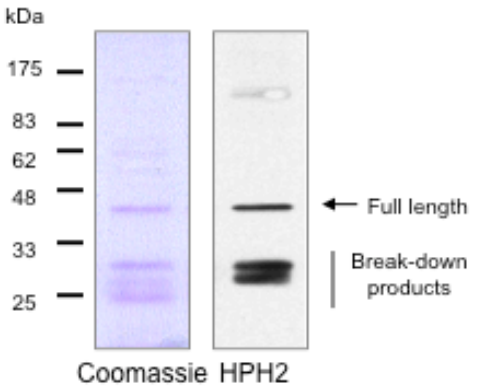
A) The cDNA of the short isoform of HPH2 (HPH2b) was fused to an N-terminal GST tag. Clones were evaluated for expression using 1mM IPTG for induction at 37°C for 4 hours or overnight at 16°C by 1mM of IPTG. A 5 l culture induced overnight at 16°C of the best clone was used for purification on glutathione sepharose 4B (GE Healthcare) and consequent cleavage of the GST tag with prescission protease (GE Healthcare). The resulting product was fractionated by SDS-PAGE and the gel was stained for both Coomassie staining and immunoblotting with a cross-reacting anti-Mph2 antibody (provided by H. Koseki)

B) HPH2b was cloned with an N-terminal HIS_6 tag. A set of clones were evaluated for expression levels following overnight induction at 16 °C by 1 mM IPTG. 2 l of the best clone were used for purification on NiNTA Superflow beads (Qiagen) and eluted using 3M imidazole as detailed in section 2.4.9.2. Although the full length protein (around 48kDa) could not be visualised by coomassie staining, it was detectable by immunoblotting for HIS and HPH2 (1615).

A

HPH2b

Molecular weight: ~48kDa (full length)
Sequence: full length
Concentration: ~0.1mg/ml
Quantity: ~0.1mg
Buffer: 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT



B

HIS$_6$-HPH2b

Molecular weight: ~44kDa (full length)
Sequence: HIS$_6$+full length
Concentration: ~0.5mg/ml
Quantity: ~0.35mg
Buffer: 3M Imidazole pH 6.0

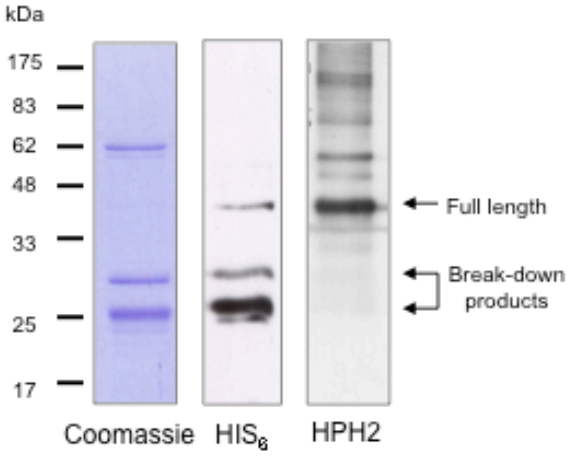


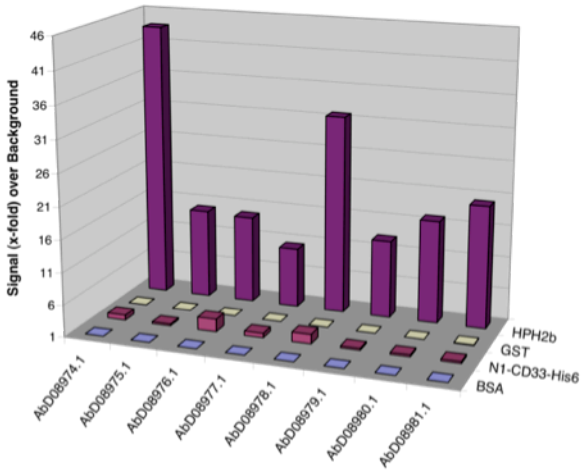
Figure 3.6 HPH2 Recombinant Human Antibodies Evaluation

A) ELISA (as performed by AbD Serotec) was used to measure the relative affinity of identified clones for the antigen HIS_6-HPH2b, as well as negative controls GST, N1-CD33-HIS_6 and BSA.

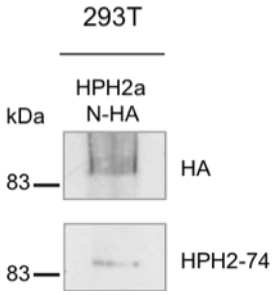
B) 293T cells overexpressing pcDNA6 HA HPH2a were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted using the 8 AbD Serotec HPH2 antibodies or anti-HA. The strongest signal, as obtained with antibody HPH2-74, is displayed.

C) 293T cells transfected with pcDNA6 FLAG-CBX7 were lysed with NP40 buffer. The lysate was then incubated with 10 μg of all AbD Serotec HPH2 antibodies and precipitated using anti-Myc (9E10) and Protein A/G agarose. The resulting immunoprecipitated samples and an input of 10% of the original lysate used (INP) were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted with anti-HA.

A


Signal (x-fold) over Background
46
41
36
31
26
21
16
11
6
1
AbD08974.1
AbD08975.1
AbD08976.1
AbD08977.1
AbD08978.1
AbD08979.1
AbD08980.1
AbD08981.1
HPH2b
GST
N1-CD33-His6
BSA

B


293T
HPH2a
N-HA
kDa
83
HA
83
HPH2-74

C

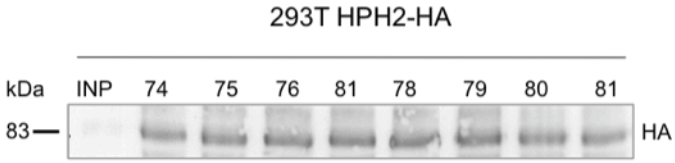
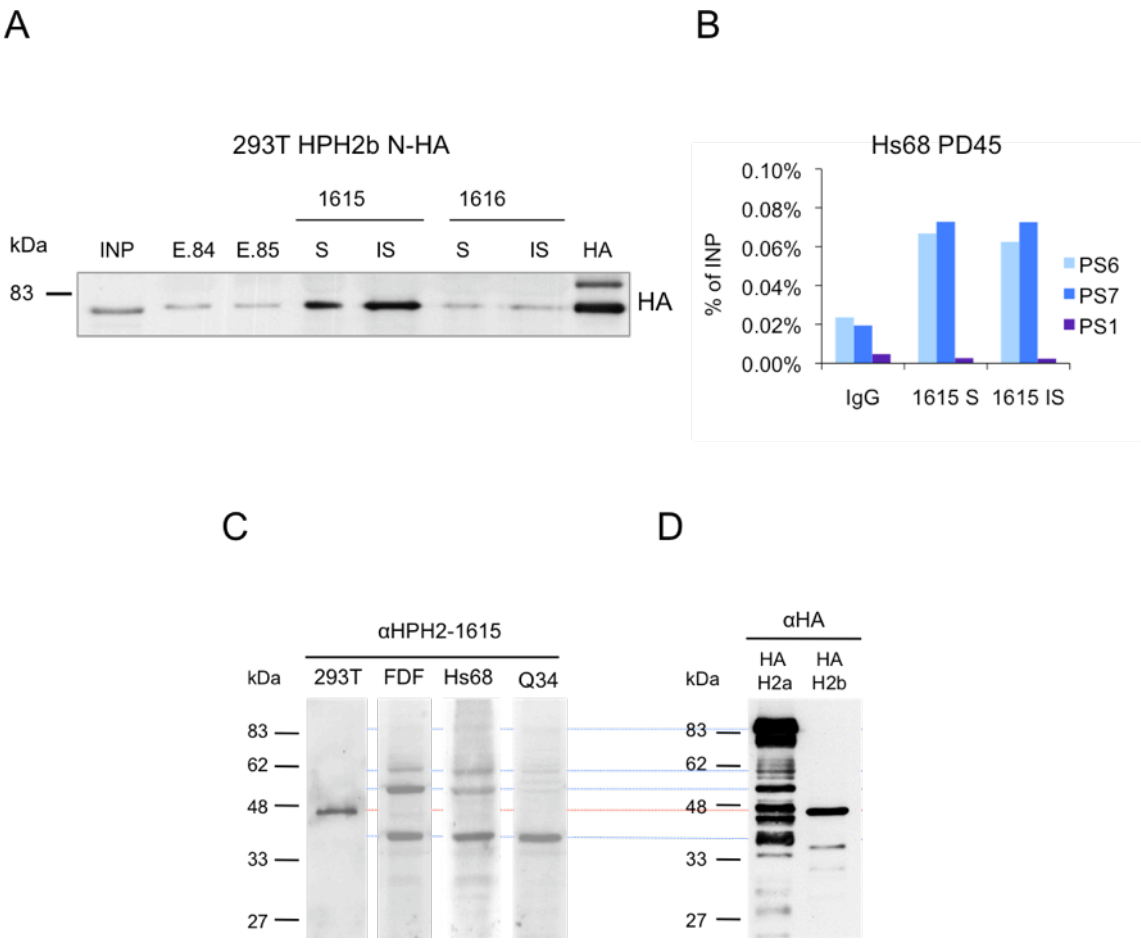

293T HPH2-HA
kDa
INP
74
75
76
81
78
79
80
81
83
HA

Figure 3.7 HPH2 Polyclonal Antibodies Evaluation.

A) 293T cells overexpressing pcDNA6 HA-tagged HPH2b were lysed with NP40 buffer. The lysates were incubated with either 5 μg of purified antibody (E.84, E.85, or anti-HA) or 20 μl of crude serum (S) or antibodies purified on Protein A (IS) (1615 or 1616) overnight at 4°C. Immunoprecipitated proteins were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted for HA. It is unclear what the additional band in the HA control IP is.

B) Chromatin Immunoprecipitation (ChIP) was performed, as described in section 2.4.5, using chromatin from mid-life Hs68 cells (PD45). The chromatin was incubated with 20 μl of crude serum (S) or antibodies purified on Protein A (IS) or unspecific rabbit IgG. The purified DNA was analysed by qPCR using primers specific to the *INK4A* promoter region (PS6), to *INK4A* exon 1 (PS7) or to *INK4B* exon 1 (PS1).

C+D) Wildtype 293T, FDF, Hs68 and Q34 cells were lysed with RIPA buffer. Protein lysates were fractionated by SDS-PAGE, transferred to nitrocellulose and immunblotted with the HPH2 1615 crude serum in a 1:500 dilution (C). The resulting autoradiographs were aligned to one based on immunoblotting the lysate from 293T cells overexpressing either pcDNA6 HA-tagged HPH2a and -HPH2b with anti-HA (D). The blue dotted lines align the sizes of bands observed in HDFs to the bands observed after ectopic expression of (C-terminal) HA-tagged HPH2a. The red dotted line aligns the size of the single band observed in 293T and the main band observed for ectopic expression of HA-tagged HPH2b.


A
293T HPH2b N-HA
1615
1616
kDa
INP
E.84
E.85
S
IS
S
IS
HA
83
HA
B
Hs68 PD45
0.10%
0.08%
0.06%
0.04%
0.02%
0.00%
% of INP
PS6
PS7
PS1
IgG
1615 S
1615 IS
C
αHPH2-1615
kDa
293T
FDF
Hs68
Q34
83
62
48
33
27
D
αHA
HA H2a
HA H2b
kDa
83
62
48
33
27

4 The Characterisation of the Polyhomeotic Family and the Interaction with the Human PC Family

4.1 Introduction

Following the identification of CBX7 as a repressor of *INK4/ARF* transcription, the question arose what its partners are in the PRC1 complex and whether these partners would potentially constitute a unique complex for the regulation of this particular locus (Gil et al., 2004; Gil and Peters, 2006). For this reason, James Nicholls, a former graduate student in our group, purified mouse Cbx7 from 293T cells by tandem affinity purification (TAP) (Nicholls, 2006). Three members of the PSC family and both members of the SCE family were found. In contrast, only one PH family member, HPH2, was found (Figure 4.1A). In subsequent *in vitro* binding experiments James also found that out of the three HPH proteins, only HPH2 was found to bind to recombinant mouse Cbx7 (Figure 4.1B).

These results collectively suggested a specific interaction between Cbx7 and only one PH member, HPH2. This was of particular interest due to the fact that Mph2 has been shown to be involved in the regulation of *Ink4a/Arf* in MEFs (Isono et al., 2005a). In this chapter the putative specific interaction between HPH2 and CBX7 was further studied by a detailed characterization of the HPH family, further assessment of specificity between the HPH and human PC protein families and by investigation of the protein domains required for this interaction.

4.2 Characterization of the HPH Family

To establish whether there are preferential interactions between the human PC and PH proteins, it was necessary to have a more comprehensive picture of the complete HPH family. At the start of this study, there was limited information about the different isoforms of the HPH genes, but at the time of writing the Ensembl, NCBI and UCSC databases collectively report several splice variants for all HPH genes (Figure 4.2A). However, a number of the variants such as HPH1c and HPH3b, -c, -d and -e do not encode the highly conserved HD1, FCS and SAM functional domains at their C-

terminus. As these domains have been demonstrated to be essential for the functions of PH in the context of PRC1, these variants were not pursued (Alkema et al., 1997; Kyba and Brock, 1998b). In order to confirm the existence of the other reported variants, a set of PCR primers was designed to discriminate differential exon usage (performed in collaboration with Tomas Racek) and primer locations are indicated in Figure 4.2A. RT-PCR was then performed on RNA from various cells types. As shown in Figure 4.2B the variants HPH1a and -c, HPH2a and -d, and HPH3a were clearly detected, but HPH1b and HPH2c were not.

In contrast to HPH1 and HPH3, HPH2 has been reported to have both a long isoform (HPH2a) of 91 kDa, and a short isoform (HPH2b) of 36 kDa that retains all C-terminal functional domains (Tonkin et al., 2002). By analogy to the mouse homologues, the shorter isoform is likely to be encoded by a distinct RNA transcript (Yamaki et al., 2002). This is consistent with the observation that the start of the 5' UTR of the transcript originates from an intronic region between exons 7 and 8 of the HPH2a transcript and thereby may have its own promoter (see Figure 4.2A). Since the cDNA of the long isoform was not available at that point, it was constructed using RT-PCR using template RNA from 293T cells. The resulting sequence corresponds to the splice variant a, and is thus missing the 5^{th} exon found in variant HPH2d.

In order to study the relative expression levels of the different HPH genes, quantitative PCR (qPCR) was used. In order to compensate for differences in primer efficiency, a standard curve derived from a titration of respective cDNA plasmid was used to determine the absolute number of transcripts for each HPH gene. HPH2b could not be specifically detected, as the high GC content of the unique 5' UTR region limits efficient PCR amplification. Upon comparison, it was found that in 293T cells all genes are expressed at a similar level, with HPH1 and HPH3 expressed $\geq$ 2-fold more than either HPH2a/d specifically or all HPH2 transcript variants (Figure 4.2C). This result also suggests that the long transcripts of HPH2 represent the majority of HPH2 transcripts in 293T cells, which is in contrast to the relative levels observed in murine embryonic and adult tissue (Yamaki et al., 2002), as ell as the findings for endogenous HPH2 in 293T, as reported in Figure 3.7C.

Using the cDNA for HPH2a and the cDNA HPH2b previously obtained, it was validated that both isoforms bind to CBX7 in a co-expression experiment in 293T cells (Figure 4.3). This suggests that, at least in the interaction with CBX7, there is little or no difference between the two isoforms. This however does not explain why full-length HPH2a or its breakdown products of more than 36 kDa were not identified in the Cbx7

purification (Figure 4.1A). This could also be due to insufficient sensitivity of the mass spectrometry analysis, especially since the more sensitive analysis performed on protein samples from tandem affinity purifications of both BMI1 and MEL18 did yield several identifications of HPH2a and its breakdown products (Maertens et al., 2009).

In the remainder of this work, HPH2a will generally be used as the representative of HPH2, as it contains additional residues with respect to the coding region of HPH2b and is more similar in size to HPH1 and HPH3. Thus HPH2 will indicate the use of HPH2a, unless specified otherwise.

4.3 Characterization of the interaction of the HPH family with the CBX family

4.3.1 All HPH family members bind CBX7

Due to unavailability of the appropriate antibodies at the start of this work, interactions were determined using FLAG epitope-mediated co-immunoprecipitation (co-IP) following ectopic co-expression in 293T cells. Each HPH family member was initially cloned with either an N- or C-terminal HA-tag to verify that the epitope would not interfere with possible interactions with the terminal regions of the protein. The epitope on either terminus of HPH2a or HPH2b did not interfere with the interaction with CBX7 (data not shown) and HPH cDNAs were expressed with an N-terminal HA-tagged tag in all subsequent experiments. Following expression of all HPH constructs in 293T cells, it was found that the proteins did not express at equivalent levels. Both HPH2 isoforms were expressed at a high level, whereas HPH1 and HPH3 could hardly be detected, as determined by immunoblotting for the HA epitope (data not shown). The inefficiency of ectopic HPH1 expression is in accordance with the findings of other investigators (Schwermann et al., 2009). To circumvent this problem, cells were transfected with increasing amounts of HPH1 and HPH3 and decreasing amounts of HPH2a/b in order to obtain equivalent levels of expression. As illustrated in Figure 4.4, it was necessary to transfect 30 µg of HPH1 to match the expression level achieved by 1 µg of HPH2a. Similar results were obtained for HPH3 (data not shown). With similar expression levels for all constructs, it was determined that all HPH proteins can co-IP

with FLAG-tagged CBX7 (Figure 4.4B). These data therefore suggest that HPH1, -2 and -3 are capable of binding to CBX7 under these conditions.

The subsequent availability of a commercial HPH1 ploYclonal antibody and a published HPH3 polyclonal antibody (Deshpande et al., 2007), in addition to the 1615 antibody described in section 3.3.2, made it possible to study the interaction between endogenous HPH proteins and CBX7. The prostate cancer cell line DU-145 was employed for this purpose as these cells express substantial levels of endogenous CBX7. DU-145 cell lysate was used for a batch immunoprecipitation of CBX7, which was divided and used for immunoblotting with either the HPH1, -2 or -3 antibody. HPH2b could not be studied by this means due to the interference of the anti-CBX7 rabbit IgG heavy chain proteins in immunoblotting. All HPH proteins were found to modestly interact with CBX7 (Figure 4.5A). In a reciprocal setup, HPH2 was able to specifically co-immunoprecipitate CBX7 (Figure 4.5B). These data therefore indicate that all HPH proteins can bind to CBX7, but that there appears to be a relatively stronger interaction between CBX7 and HPH2a, which may indicate a preference of interaction.

To investigate the interaction between HPH2 and CBX7 in more detail, a series of deletion mutants were generated of HA-HPH2 to sequentially remove the known C-terminal functional domains (Figure 4.6A). These mutants were then tested for their ability to precipitate with FLAG-tagged CBX7 following ectopic co-expression in 293T cells. These co-immunoprecipitations demonstrate that the FCS domain is required for the interaction with CBX7 (Figure 4.6B). Since the FCS domain is highly conserved between the HPH family proteins, particularly at its key cysteine residues, it can be deduced that the FCS domain could be responsible for the interaction of CBX7 with the HPH proteins.

4.4 All human PC proteins bind to HPH2

Following the finding that CBX7 interacts with all HPH proteins, it was of interest to determine whether there is any inverse specificity in the association of HPH2 with the other human PC proteins. To address this question, 293T cells were co-transfected with plasmids encoding HA-tagged HPH2a and FLAG-tagged CBX2, CBX4, CBX6, CBX7 and CBX8 (kindly provided by Marc Rodriguez-Niedenführ). The cell lysates were then used for immunoprecipitation of the FLAG epitope and precipitates were

immunoblotted for HA. As shown in Figure 4.7, HPH2 appeared to interact with all these CBX proteins under these conditions. An obvious inference is that the interaction involves a domain conserved in all CBX proteins.

Pc/Cbx/CBX proteins contain an amino terminal chromodomain, through which they interact with histones and RNA (Bernstein et al., 2006b). A number of key residues have been characterised, which, upon mutation, can abrogate these interactions. The F11 and W35 residues have been shown to mediate binding of the chromodomain to the trimethylated lysine 27 on the tail of histone H3 (Bernstein et al., 2006b), whereas the R17 residue is believed to be required for RNA binding (M Walsh, personal communication). It was therefore of interest to determine whether any of these mutations would disturb the interaction between CBX7 and HPH2. Versions of FLAG-tagged CBX7 carrying these mutations (kindly provided by Marc Rodriguez-Niedenführ) were co-expressed with HA-tagged HPH2 in 293T cells and the interaction was assessed as before. Figure 4.8 shows that none of these mutations affected the interaction of CBX7 with HPH2, suggesting that these residues may not be involved in the interaction and that the interaction may occur independently of these functions of CBX7.

In order to dissect the domain of CBX7 involved in the interaction with the HPH proteins, a series of seven FLAG-tagged CBX7 deletion mutants was made. This series was in part based on the CBX7 fragments previously cloned for bacterial expression (discussed in section 3.2.1). In contrast to full length CBX7, none of the CBX7 deletions could be expressed at detectable level following transient transfection of 293T cells. In addition, *in vitro* transcription and translation of these constructs did not yield any detectable products, again in contrast to full length CBX7. These results could be due to the instability of human CBX7, as previously discussed in Chapter 3.

As an alternative strategy, a GST pulldown assay was used. The mutants were cloned as N-terminal GST-fusion products and were tested for bacterial expression. Three out of the seven deletion mutants constructed could be expressed well, including the two CBX7 GST-fusion proteins described in Chapter 3. Of three mutants, only two were found to be stably recovered following the GST pulldown procedure; Δ170C (Δ1) and Δ47CD-Δ170C (Δ2) (Figure 4.9A). These 2 mutants, GST-Cbx7 and GST were subsequently used in GST pulldown experiments with *in vitro* transcribed and translated ^{35}S-methionine labelled HPH proteins and luciferase, used as a negative control. As shown in Figure 4.9B, Δ1 most efficiently pulled down all HPH proteins. Δ2 did not consistently pull down all HPH proteins and overall did so with lower efficiency (with

the exception of HPH1 pulldown). These two mutants both include amino acids 48-81 (aa48-81) and thereby suggest that these residues are mediating the interaction. The aa48-81 region of CBX7 corresponds to the distal portion of the chromodomain and is highly conserved in the human CBX proteins, all of which are able to interact with HPH2 (see Figure 4.7).

However, there are a number of unsettling aspects of the GST pulldown data. Firstly, the GST negative control produced a signal for both HPH2a and HPH2b. This could in part be attributed to apparent GST overloading (HPH2b) and could overall be resolved by increased stringency in washing conditions. Secondly, HPH1, HPH2a and HPH3 were not found to bind mouse Cbx7. This is consistent with previous GST pulldown data (Figure 4.1B), but does not fit with the observations that HPH1, HPH2a and HPH3 clearly interact with human CBX7 following co-expression in 293T cells (Figure 4.4). This may be explained by the conformational constraints under these artificial conditions, which may differently affect the interaction the smaller-sized HPH2b. In order to connect these results to the co-immunoprecipitation of all HPH proteins by CBX7 (Figure 4.4), the assay should ideally have been performed with human CBX7, but this cannot be recombinantly expressed as discussed before (Chapter 3). Noting these reservations, this data suggests that the aa48-81 region of CBX7 is responsible for the interaction with HPH proteins and its high degree of conservation may explain why all HPH proteins can interact with all CBX proteins.

4.5 Discussion

In summary, the presented data did not demonstrate evidence for binding specificity between the CBX and HPH family members. The presented results suggest that there is no structural specificity in the interaction of HPH and CBX proteins, that there is a direct interaction between these protein families and that this interaction occurs through highly conserved domains. Nevertheless, endogenous protein interactions hint at possible preferential binding between CBX7 and HPH2.

Practical limitations therefore might best explain the observations from James Nicholls, as presented at the start of this chapter. The low sensitivity of the mass spectrometry analysis of the protein samples obtained from the purification of mouse Cbx7 presents an important constraint in evaluating possible specificity of interactions. All proteins identified by mass spectrometry were represented by a small number of peptides and for HPH2 specifically, not more than two peptides were found in one purification experiment. The analysis thus may not have been sufficiently sensitive to identify peptides corresponding to HPH2a, or to HPH1 and HPH3. In addition, the HPH proteins were detected at similar expression levels in 293T (Figure 4.1C), suggesting that the identification of only HPH2b does not correlate to the expression level of the individual HPH proteins, although protein instability as demonstrated for these proteins may be a concern. Furthermore, the conclusion from the *in vitro* binding assay of HPH proteins to GST-mouse Cbx7 as performed by James Nicholls (Figure 4.1B) does not appear to be correct.

The co-precipitation of all HPH proteins with CBX7 following co-expression in cells, and *in vitro* binding to a specific domain of CBX7 suggests that all HPH proteins can directly interact with CBX7 under these experimental conditions. It is however important to recognise the limitations of these techniques. Overexpression of tagged proteins and the immunoprecipitation of a foreign epitope can lead to upnspecific or high background binding. One major drawback in the data presented for these experiments is the absence of a control sample in which the protein targeted for precipitation is absent and hence will reveal whether the protein studied is indeed unable to bind without it. In addition, the use of recombinant fragments of protein has its limitations. Since only a fraction of the protein is used, it is questionable to what extent the native structural confirmation is maintained and therefore to what extent a native interaction can be reproduced. This concern is strengthened by the observation

that HPH proteins do not precipitate strongly with GST-mCbx7, even though it contains highly conserved aa48-81 region. As the GST pulldown experiments were controlled for unspecific GST binding, it is unlikely that there were serious problems with background binding. Considering these reservations, we can conclude that the data demonstrates a putative direct interaction between the HPH and CBX proteins. It is however not straightforward to indisputably validate a direct interaction.

Furthermore, a direct interaction between CBX and HPH proteins would go against previous findings in *Drosophila*, where Ph was not found to interact with Pc (Kyba and Brock, 1998a). It is interesting to speculate on the role of this interaction in the context of the PRC1 complex. The interaction between HPH and PSC has been well established, Ph was demonstrated to bind directly to the helix-turn-helix-containing (HTH) region of Psc through its HD1 domain (Kyba and Brock, 1998a). Additional evidence from studies with mouse constructs found that both the HD1 and SAM domain of Mph1 were required for the interaction with Bmi1 and this was confirmed for human proteins using both HPH1 and HPH2 (Alkema et al., 1997; Gunster et al., 1997). Furthermore, our laboratory has been able to co-purify all HPH proteins with BMI1 or MEL18. RING1A and Ring1b have also both been demonstrated to interact with HPH1 and Mph1 respectively (Satijn et al., 1997; Suzuki et al., 2002), but it is not sure if this would involve a direct interaction. It may thus be that the HPH proteins have direct interactions with all core PRC1 components.

The putative direct CBX-HPH interaction would suggest that the status of HPH2 can indeed affect the function of CBX7 in the repression of *INK4/ARF*. However, since these data have ruled out the existence of inherent specificity between these two proteins, this role may not only be confined to HPH2 and suggests that HPH1 and HPH3 should also be considered further. Nevertheless, it remains to be determined which of these possible interactions between the CBX and HPH family (15, disregarding the different isoforms or splice forms of HPH2) occur in what context and to what degree there is redundancy in function.

Figure 4.1 Potential specificity of the interaction between CBX7 and HPH2b (adapted from (Nicholls, 2006))

A) Total cell lysate from 1.2 x 10^9 of 293T cells stably expressing either TAP or mouse Cbx7-TAP was used to carry out a large-scale TAP-tag purification. The resulting protein sample was separated by SDS-PAGE and the gel was stained with Brilliant Blue G (Sigma) O/N. Visible bands were excised and subjected to tryptic digestion. Bands observed in the control lane were not submitted for mass spectrometry analysis. Peptides were identified as described in the Materials and Methods section. PRC1 proteins found to co-purify with mouse Cbx7 are labelled with a red asterisk.

B) 1 μg of plasmid DNA encoding HPH1, HPH2 or HPH3 was used for coupled *in vitro* transcription and translation in the presence of ^{35}S-methionine using the TnT kit (Promega). *0.5* μl HPH1, 7.5 μl HPH2 and 0.5 μl HPH3 were incubated with 20μg of GST or GST-Cbx7 pre-bound to blocked glutathione beads. Proteins bound to the beads were then fractionated by SDS-PAGE alongside an equivalent volume of *in vitro* translated material (IVT) used in the binding assay. Radiolabelled proteins were then detected by autoradiography.

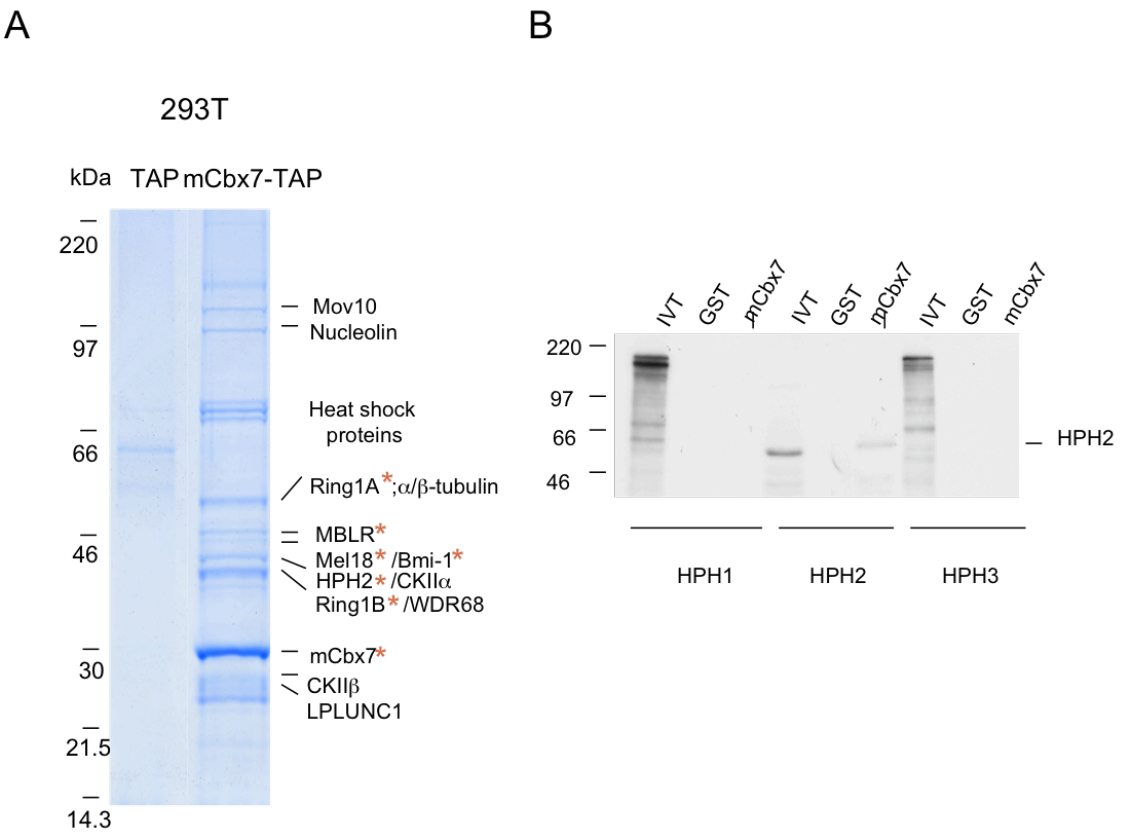

A
293T
kDa
TAP
mCbx7-TAP
220
97
66
46
30
21.5
14.3
Mov10
Nucleolin
Heat shock proteins
Ring1A*;α/β-tubulin
MBLR*
Mel18*/Bmi-1*
HPH2*/CKIIα
Ring1B*/WDR68
mCbx7*
CKIIβ
LPLUNC1
B
IVT
GST
mCbx7
IVT
GST
mCbx7
IVT
GST
mCbx7
220
97
66
46
HPH2
HPH1
HPH2
HPH3

Figure 4.2 The characterisation of HPH transcripts in human cells

A) Overview of human polyhomeotic (HPH) members including a simplified cDNA structure, highlighting the location of the conserved C-terminal domains, and annotated transcripts as collectively listed in the Ensembl, UCSC and NCBI databases. PCR primer locations for detecting alternative usage of exons of interest are indicated by arrowheads aligned with the relevant exons. Sequences are detailed in Table 6.

B) Discriminative PCRs for the validation of annotated splice variants of HPH1, HPH2 and HPH3 were performed using the standard conditions (as detailed in the Materials and Methods section 2.2.2) with 0.4 μg of cDNA of 293T and U2OS cells (HPH1 and HPH3 variants) or cDNA of O2Os and H1299 (HPH2 variants) cells. PCR reactions were separated by agarose gel electrophoresis and visualised by exposing the gel to UV following ethidium bromide staining.

C) Quantitative RT-PCR was performed using 0.4 μg of RNA from 293T cells relative to specific plasmid standard curves to determine transcript copy number of HPH1, HPH2, HPH3 and the HPH2a/d variants specifically. Error bars indicate standard deviation between biological replicates, with n = 3.

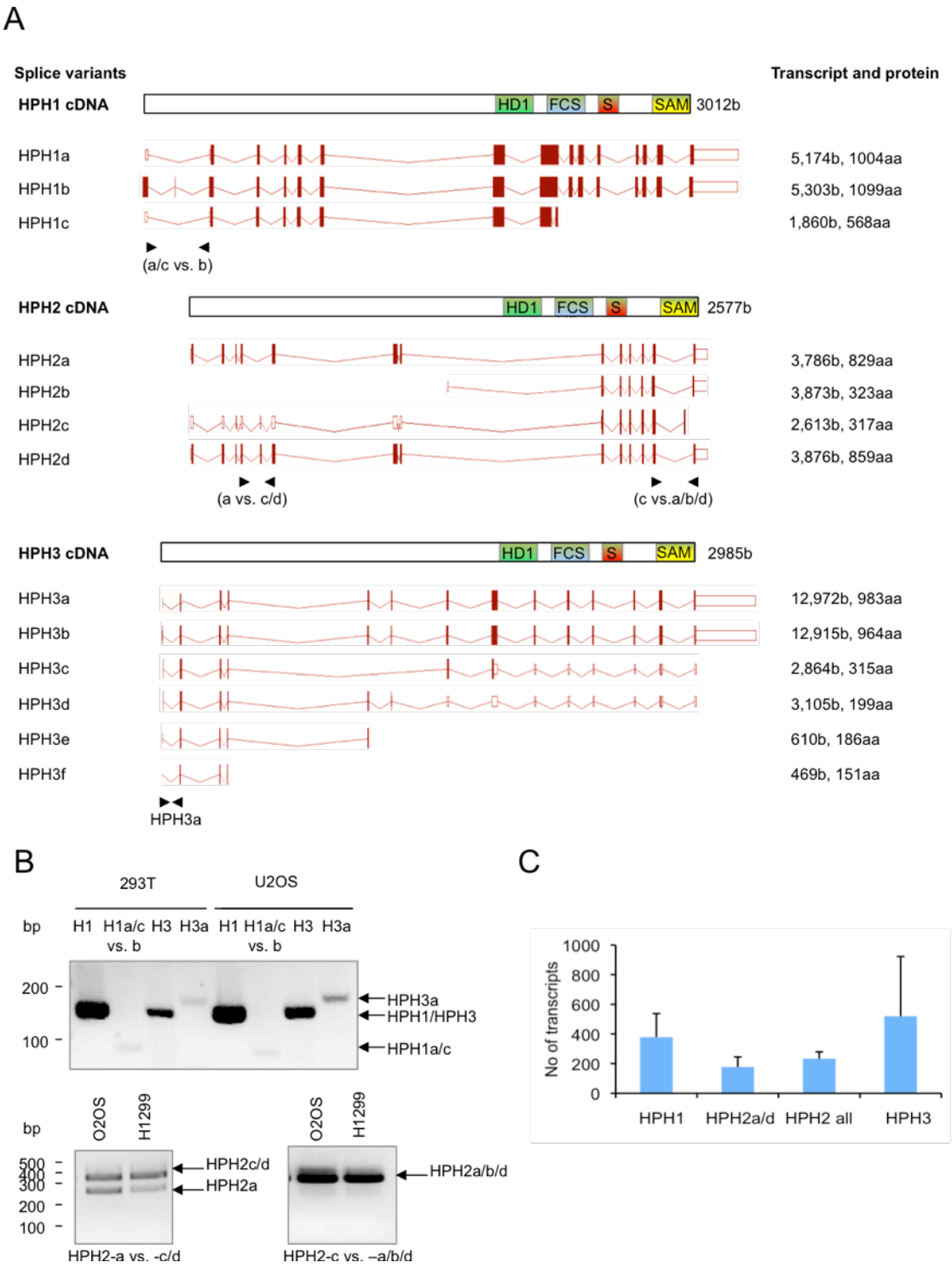

A
Splice variants
Transcript and protein
HPH1 cDNA
HD1
FCS
S
SAM
3012b
HPH1a
HPH1b
HPH1c
5,174b, 1004aa
5,303b, 1099aa
1,860b, 568aa
(a/c vs. b)
HPH2 cDNA
2577b
HPH2a
HPH2b
HPH2c
HPH2d
3,786b, 829aa
3,873b, 323aa
2,613b, 317aa
3,876b, 859aa
(a vs. c/d)
(c vs.a/b/d)
HPH3 cDNA
2985b
HPH3a
HPH3b
HPH3c
HPH3d
HPH3e
HPH3f
12,972b, 983aa
12,915b, 964aa
2,864b, 315aa
3,105b, 199aa
610b, 186aa
469b, 151aa
HPH3a
B
293T
U2OS
bp
H1
H1a/c vs. b
H3
H3a
200
100
HPH3a
HPH1/HPH3
HPH1a/c
O2OS
H1299
500
400
300
200
100
HPH2c/d
HPH2a
HPH2a/b/d
HPH2-a vs. -c/d
HPH2-c vs. –a/b/d
C
No of transcripts
1000
800
600
400
200
0
HPH1
HPH2a/d
HPH2 all
HPH3

Figure 4.3. CBX7 interacts with both isoforms of HPH2.

293T cells were co-transfected with pcDNA6 FLAG CBX7 and pcDNA6 encoding HA–tagged HPH2a or -HPH2b. Whole cell lysates were subjected to immunoprecipitation (IP) using the anti-FLAG antibody (as described in section 2.4.4). Purified proteins (IP) and 5% input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA or anti-FLAG.

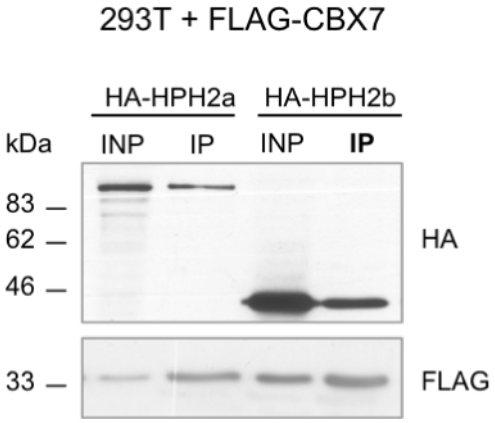


Figure 4.4. CBX7 interacts with all three HPH proteins

A) 293T cells were transfected with a titrated amount of pcDNA6 HA HPH2a (1-4μg) or HA HPH1 (10-50 μg). Whole cell lysates were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA or anti-CDK4.

B) 293T cells were co-transfected with pcDNA6 FLAG CBX7 and 30μg of HPH1, 1 μg of HPH2a and 30 μg of HPH3 in a pcDNA6 HA construct. Whole cell lysates were subjected to immunoprecipitation (IP) using the anti-FLAG antibody (as described section 2.4.4). Purified proteins and 10% of input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA (for horizontal lanes labelled HPH1, HPH2, HPH3) or anti-FLAG (for lane labelled FLAG). The FLAG blot displayed is a representative of all IPs.

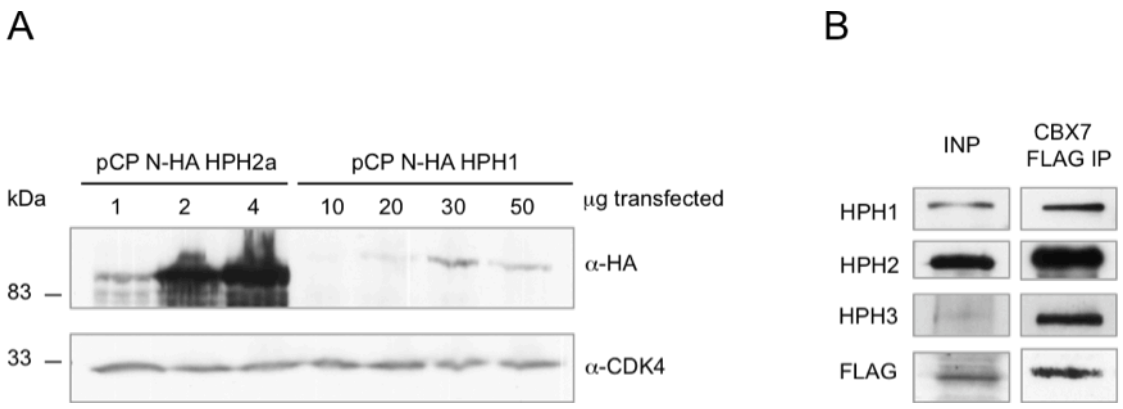


Figure 4.5 Preferential binding of HPH2 to endogenous CBX7.

A) DU-145 cells were lysed and subjected to immunoprecipitation (IP) using anti-Cbx7 (Abcam). Purified proteins (CBX7 IP) and 10% input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HPH1 (Abnova), anti-HPH2a (1615) and anti-HPH3 (obtained from M.F. Hansen).

B) DU-145 cells were used for IP as in A), using anti-HPH2 (1615) and an unspecific IgG fraction as a mock control. Immunoblotting of precipitates (HPH2 IP) and input (INP) was performed as in A), with anti-CBX7 (Abcam).

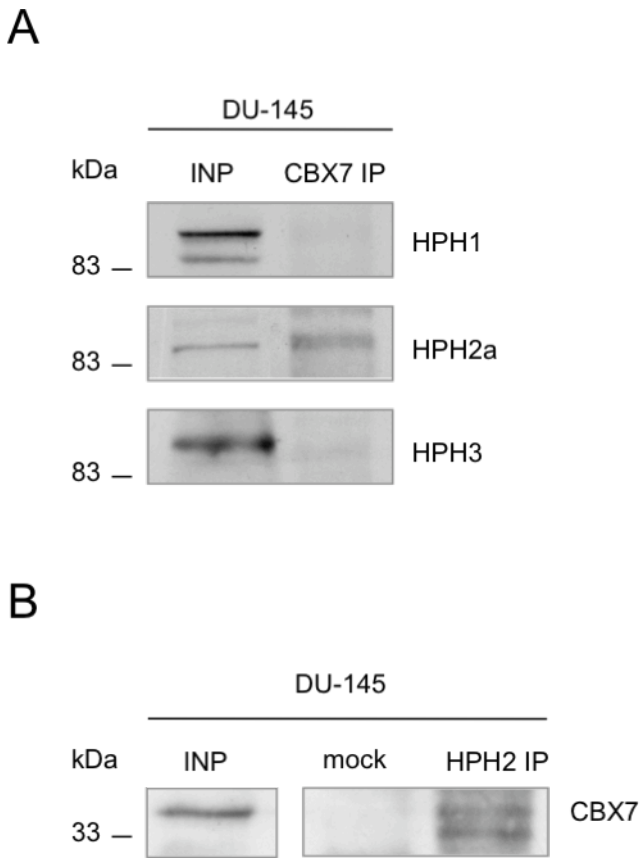


Figure 4.6 CBX7 interacts with the FCS domain of HPH2

A) Overview of the HA-tagged HPH2a deletion mutants, displaying the cumulative deletion of C-terminal functional domains.

B) 293T cells were co-transfected with pcDNA6 FLAG-CBX7 and pcDNA6 encoding HA-HPH2a wildtype, -ΔSAM, -ΔSAMΔpS or -ΔSAMΔpSΔFCS, or empty vector control. Whole cell lysates were subjected to immunoprecipitation (IP) with the anti-FLAG antibody (as described in section 2.4.4). Purified proteins (FLAG IP) and 5% input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA or anti-FLAG.

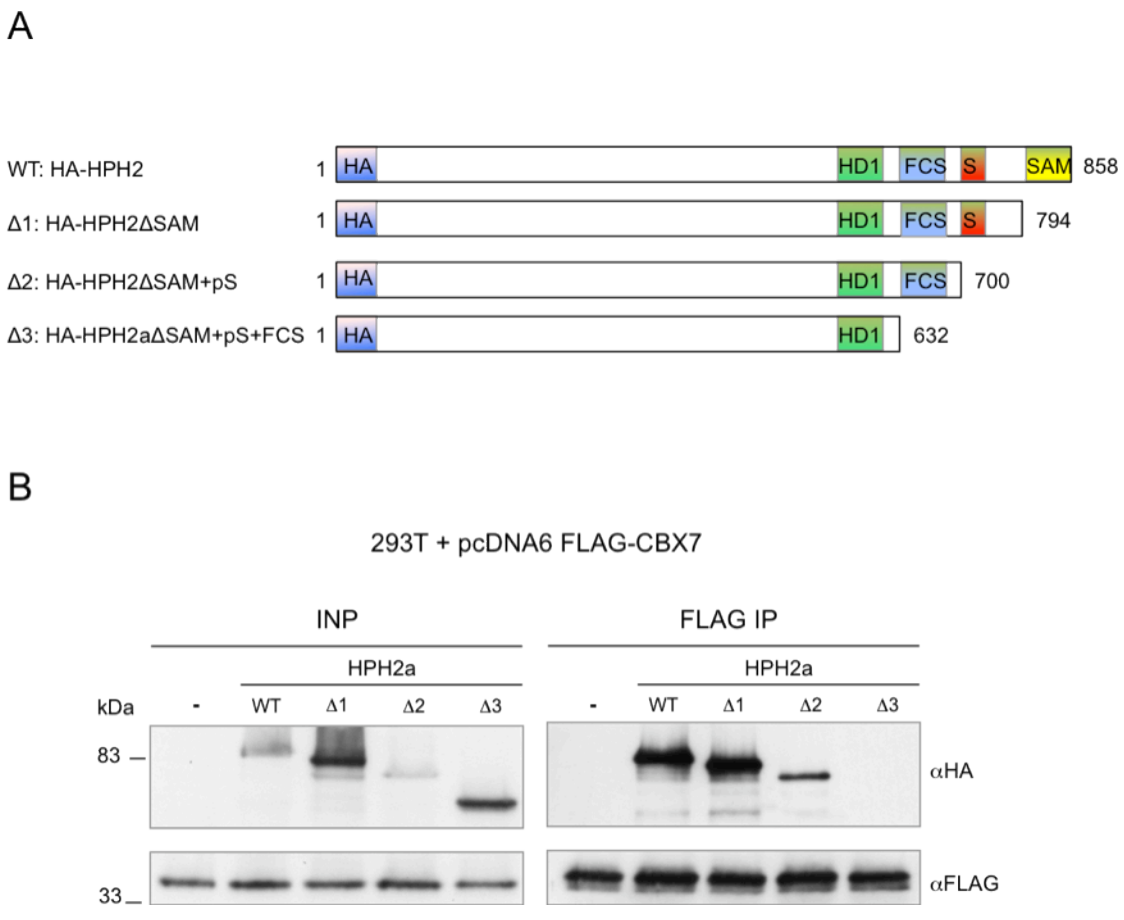


Figure 4.7 HPH2 interacts with all five human PC proteins.

293T cells were co-transfected with pcDNA6-based constructs encoding FLAG-tagged CBX2, CBX4, CBX6, CBX7 or CBX8, and HA-tagged HPH2. Whole cell lysates were subjected to immunoprecipitation (IP) with the anti-FLAG antibody (as described in 2.4.4). Purified proteins (FLAG IP) and 5% input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA or anti-FLAG.

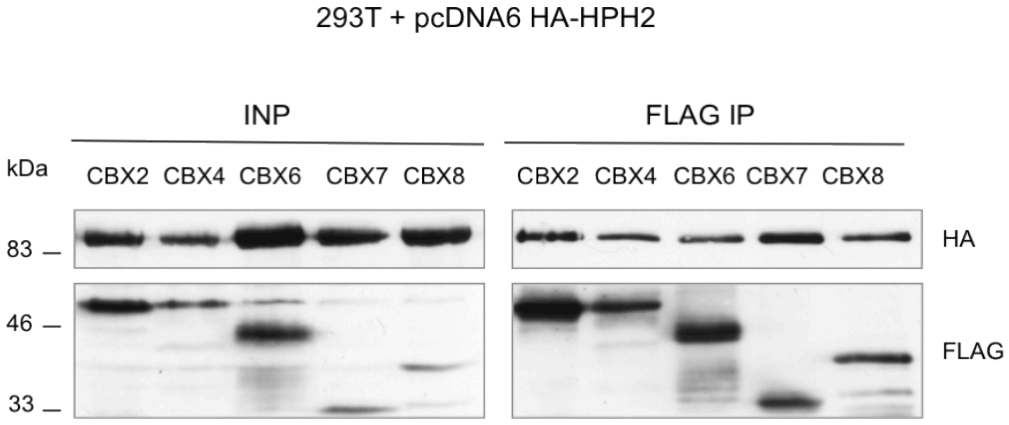


Figure 4.8 Chromodomain mutations do not affect the interaction of CBX7 with HPH2

293T cells were co-transfected with pcDNA6 plasmids encoding HA-tagged HPH2, and FLAG-tagged CBX7 wildtype or F11A, W35A or R17Q mutants. Whole cell lysates were subjected to immunoprecipitation (IP) for the FLAG-epitope (as described in section 2.4.4). Purified proteins (FLAG IP) and 2% input samples (INP) were fractionated by SDS-PAGE, transferred to PVDF and immunoblotted with anti-HA or anti-FLAG.

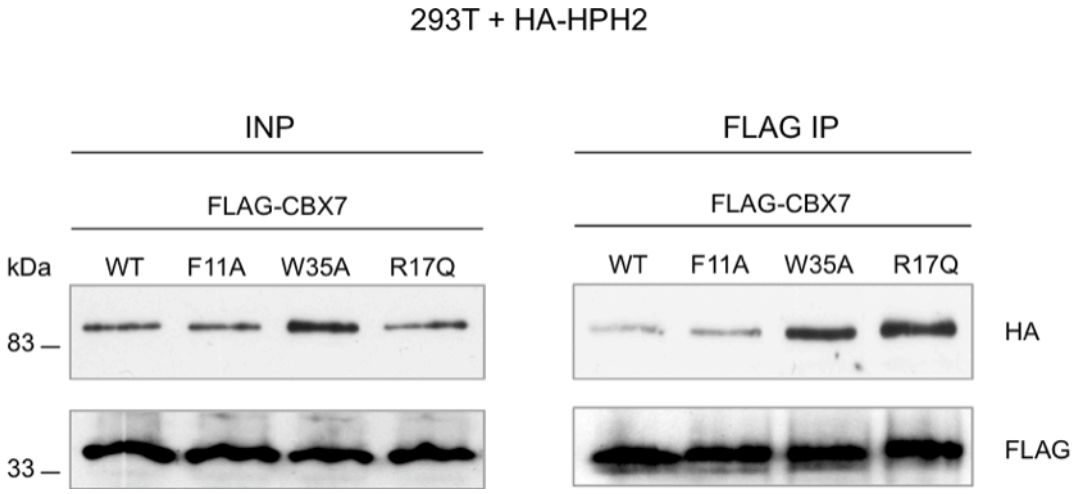


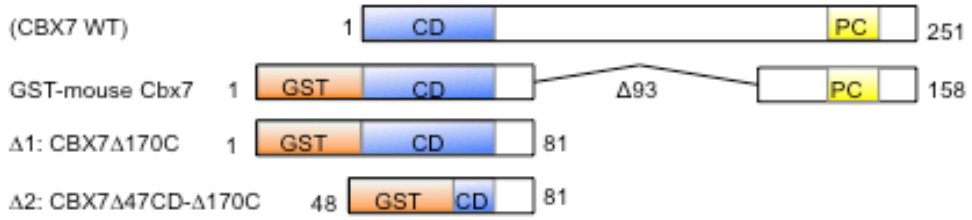
Figure 4.9 HPH proteins interact with a highly conserved domain of CBX7

A) Overview of the CBX7 deletions produced as GST-fusion proteins (with respect to wildtype), which include a combination of N- and C-terminal truncations. The region indicated by Δ93 is the region that is missing in the mouse relative to the human sequence.

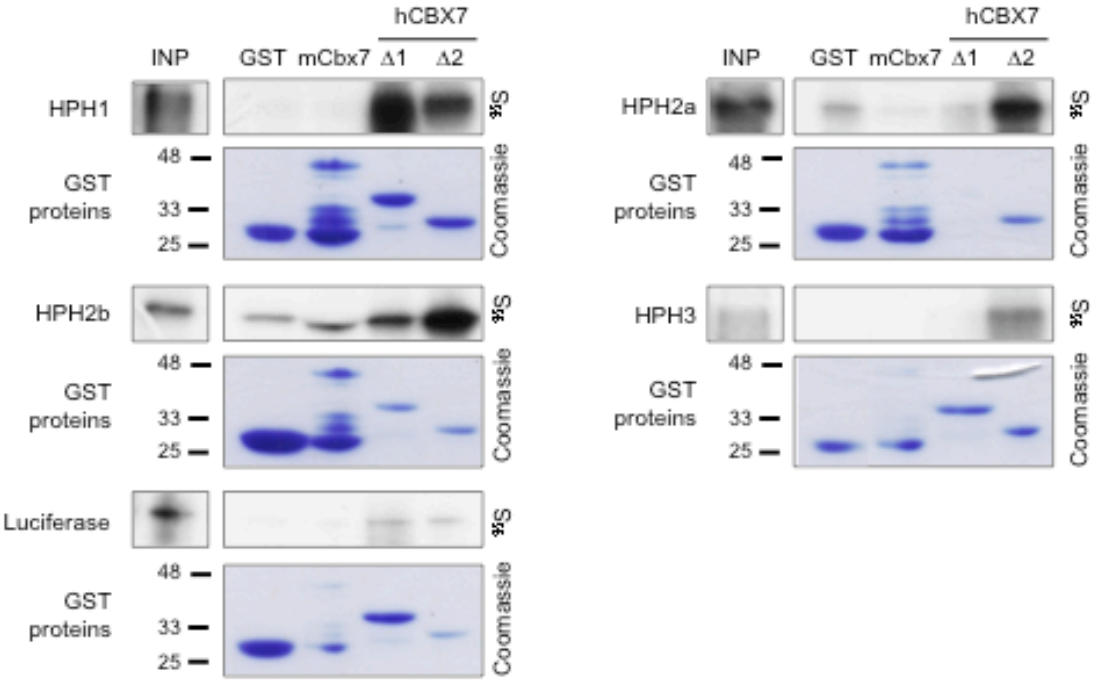
B) 1μg of plasmid DNA encoding HPH1, HPH2a, HPH2b, HPH3 or luciferase was used for coupled *in vitro* transciption and translation in the presence of ^{35}S-methionine using the TnT kit (Promega). 8 μl of the 50 μl reaction was used for incubation with 20 μg of GST, GST-Cbx7, GST-CBX7Δ170C or GST-CBX7Δ47CD+Δ170C pre-bound to blocked glutathione beads. Proteins bound to the beads were then fractionated by SDS-PAGE alongside 10% input samples (INP) of *in vitro* translate. Radiolabelled proteins were then detected by autoradiography and the GST pulldown efficiency was checked by Coomassie staining

C) Overview of sequence conservation of aa48-81 of CBX7 relative to all other human PC proteins and mouse Cbx7, with the aa48-81 region and corresponding perfectly conserved residues highlighted in yellow.

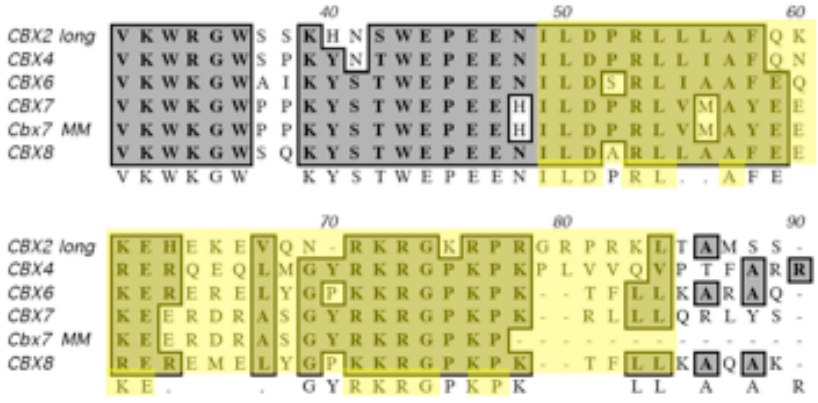
A


(CBX7 WT)
1
CD
PC
251
GST-mouse Cbx7
1
GST
CD
Δ93
PC
158
Δ1: CBX7Δ170C
1
GST
CD
81
Δ2: CBX7Δ47CD-Δ170C
48
GST
CD
81

B


hCBX7
INP
GST
mCbx7
Δ1
Δ2
HPH1
35S
GST proteins
48
33
25
Coomassie
HPH2b
HPH2a
HPH3
Luciferase

C


CBX2 long
CBX4
CBX6
CBX7
Cbx7 MM
CBX8

5 The Role of the Polyhomeotic Family in the Regulation of *INK4A*

5.1 Introduction

The results in Chapter 2 suggest that all three members of the human polyhomeotic (HPH) family can physically associate with CBX7, but did not address functional interdependence. In mouse knockouts studies, only the murine homolog of HPH2 was clearly linked to the regulation of *Ink4a* (Isono et al., 2005a). It was therefore important to extend the analysis to all HPH proteins to define whether there is a collective involvement of the HPH family in the regulation of the *INK4/ARF* locus. To this end, multiple shRNAs were designed and validated for depletion of HPH1, -2 and -3 expression. Conversely, the respective cDNAs were cloned into retroviral vectors to achieve overexpression in primary cells. However, the initial focus was on HPH2, based on the findings of the corresponding mouse knockout, the indications of some degree of preferential binding by endogenous CBX7 and the availability of reagents at the start of this work.

By definition, all established human cell lines have alterations that allow them to bypass senescence. As most of these reflect loss or deregulation of *INK4A/ARF*, it is essential to study the regulation of the locus in primary cells. The system of choice in the laboratory is primary human fibroblasts of various origins that have been transduced with the receptor for ecotropic retroviruses (Brookes et al., 2004). Unfortunately, primary cells are rather heterogeneous in nature and therefore any effects on *INK4A* expression are averaged out over the population sampled. For instance, it has been observed that at mid-age a population of primary fibroblasts will already have a fraction of senescent cells with high levels of p16 (Brookes et al., 2004). This heterogeneity can therefore obscure the effects that one is aiming to document from an experimentally manipulated population.

Due to the overall increase in the percentage of p16-expressing cells in culture over time, it is important to consider the age or the number of population doublings the cells have undergone for experimental purposes. As a population of older cells has a higher overall basal level of p16 protein, these cells were used for documenting repression of *INK4A*. Conversely, a population of younger cells has a lower overall

level of basal p16 protein and was therefore preferred for observing derepression of *INK4A*. For this reason, the number of population doublings for the cells used for each experiment was chosen carefully and was stipulated as appropriate. Most of the experiments were conducted with neonatal foreskin fibroblasts (Hs68) or fetal dermal fibroblasts (FDF), the characteristics of which have been described (Brookes et al., 2004).

5.2 HPH2 knockdown in HDFs results in derepression of INK4a

Initially, an effort was made to validate whether the effects on *Ink4a/Arf* derepression as reported for the Mph2-/- knockout MEFs, also apply to human cells (Isono et al., 2005a). To knockdown the HPH2 transcripts, a set of 6 shRNAs were designed and cloned into the retroviral vector pRetroSuper (pRS). Since at that time no antibody capable of detecting endogenous HPH2 was available, these constructs were tested for their ability to deplete HA-tagged HPH2 protein levels expressed in 293T cells (validation of best shRNA shown in Figure 5.1A). The shRNA capable of displaying the best knockdown, HPH2-1, was consequently used to infect FDF cells. It was observed that knockdown of HPH2 resulted in increased levels of p16 protein, suggesting derepression of *INK4A* (Figure 5.1B). This effect was similar to that seen for knockdown of CBX7, which was previously shown to cause an induction of *INK4A* and result in pre-mature senescence of normal human cells (Gil et al., 2004; Maertens et al., 2009). The induction of *INK4A* expression by HPH2 and CBX7 was confirmed at the RNA level (Figure 5.1C). Interestingly, it was found that HPH2 also induces the expression of *INK4B*, but not *ARF*.

To demonstrate that derepression of *INK4A* following HPH2 knockdown also impacts on proliferation, population doublings were recorded following completion of selection. The resulting graph (Figure 5.1D) demonstrates a reduction in proliferation to the point where the shRNA-transduced cells stopped dividing (at PD37, 5 population doublings after selection), whereas wildtype cells normally continue proliferating up to PD80 (Brookes et al., 2004). CBX7 knockdown was used as a positive control and reduced proliferation to a similar extent. Additionally, shRNA transduced cells (at PD37) and control cells (PD39) were stained for senescence-associated ß-galactosidase (SA ß-gal). It was found that both HPH2 and CBX7 knockdown lead to a significant increase in the intensity of SA ß-gal staining (Figure 5.1E), confirming that the arrest

corresponds to premature senescence. With the caveat that only one shRNA was used in this experiment, these results clearly demonstrate that that HPH2 is required for *INK4A* repression in human cells.

5.3 *HPH2 binds to the promoter and exon 1 region of* INK4A

If HPH2 is directly involved in PcG-mediated repression of *INK4A*, it would be expected to be associated with the chromatin at the locus. To address this possibility, chromatin immunoprecipitation (ChIP) experiments were conducted with chromatin from mid-lifespan FDF cells using the 1615 rabbit polyclonal antibody against HPH2 (as validated in section 3.3.2). Other PRC1 members BMI1, CBX7 and CBX8, previously shown to bind to *INK4A,* were used as positive controls (Dietrich et al., 2007; Kia et al., 2008; Maertens et al., 2009). A non-specific rabbit IgG or a mouse monoclonal against HA were used as negative controls. Using primer pairs spanning the *INK4B-ARF-INK4A* locus (Figure 5.2A), HPH2 was significantly enriched in the region from upstream of the promoter region to downstream of the first exon of *INK4A* (Figure 5.2B). The distribution of HPH2 is similar to that of CBX7, CBX8 and BMI1 (Figure 5.2C-E). H3K27me3, the hallmark of PRC1 binding, follows a similar, though somewhat wider distribution pattern, consistent with previous observations (Figure 5.2F) (Dietrich et al., 2007; Maertens et al., 2009). These findings correspond to the previous study in MEFs that showed Mph2 to be associated with the first and second exon of *Ink4a* (Isono et al., 2005a). These data provide a better definition of *INK4B-ARF-INK4A* locus-wide binding and confirm the suspected overlap with other PRC1 proteins at *INK4A*.

5.4 HPH2 depletion does not affect repression of INK4A by Cbx7

As previously discussed, ectopically expressed mouse Cbx7 is able to downregulate the expression of *INK4A/ARF* (Gil et al., 2004). It is however not clear how the additional availability of Cbx7 enhances repression relative to endogenous CBX7 or what PRC1 partners are involved in this process. It was therefore important to establish whether the ability of mouse Cbx7 can down-regulate the expression of *INK4A* in the face of HPH2 knockdown.

As expected, expression of FLAG-tagged mouse Cbx7 in Hs68 cells caused a marked reduction in the levels of $p16^{INK4A}$ protein and RNA at 10 days post-selection (Figure 5.3A+B). This was accompanied by increased binding of exogenous FLAG-tagged Cbx7 at the *INK4A* promoter (Figure 5.3C). The cells were then infected with retroviruses encoding either HPH2 or CBX7 shRNA. Whereas the control cells showed significant derepression of *INK4A*, the cells overexpressing mouse Cbx7 did not display such a response to depletion of HPH2 (Figure 5.3D). CBX7 knockdown did have a modest effect on relieving *INK4A* repression. It should be noted that the shRNA against CBX7 has a 2 base mismatch (out of the 18 bases) with the sequence of mouse Cbx7 and it is therefore unclear whether the shRNA will also target the ectopically expressed mouse Cbx7, which would be expected to impact on the down-regulation of *INK4A*. An additional approach using lentiviral infection with a different HPH2 shRNA produced similar results (Figure 5.3D). This would imply that repression of *INK4A* by mouse Cbx7 in these cells is not sensitive to depletion of HPH2 (nor endogenous CBX7).

It thus appears that depletion of HPH2 has little effect on the ability of mouse Cbx7 to down-regulate *INK4A* expression. One obvious possibility is that the knockdown of HPH2 was incomplete and therefore sufficient HPH2 remains present for PcG-mediated repression. Alternatively, it is possible that ectopic Cbx7 is able to produce *INK4A* repression by interacting with HPH1 and HPH3. Another scenario would be that endogenous CBX7 (in the vector control cells) is present at a rate-limiting amount for PcG repression, so that its function is very sensitive to knockdown by CBX7-1 or to knockdown of an essential partner, such as HPH2. In contrast, the ectopically expressed mouse Cbx7 would have greater potential to form alternative partnerships, for example with HPH1 and HPH3, and would therefore be less sensitive to HPH2 depletion.

5.5 **Ink4a** *repression by Cbx7 is incapacitated in Mph2$^{-/-}$ MEFs*

As mentioned above, the incomplete depletion of HPH2 might obscure any level of interdependence with Cbx7. As an alternative, we obtained *Mph2* knockout MEFs from the Koseki laboratory (Isono et al., 2005a) to test whether Cbx7 is able to repress *Ink4/Arf* in the complete absence of Mph2. Although these cells are reported to achieve 7 population doublings (compared to ~15 population doublings for wildtype MEFs), cells as received were unfortunately found to senesce after only 2-3 population doublings, allowing very little opportunity for experimental manipulation and low amounts of material for analysis.

Because of these limitations, it was first necessary to establish conditions at which the relevant retroviral constructs delivered the expected effects in wildtype MEFs. Overexpression of mouse Cbx7 or an shRNA against *Cdkn2a,* which knocks down both the Arf and p16 transcripts, caused a substantial reduction in p16 protein levels (Figure 5.4A). Consequently, the same plasmids and a plasmid containing mouse Mph2 were used to infect wildtype and *Mph2-/-* MEFs. Following selection, cells were imaged to document both cell morphology and relative cell number and their proliferation was followed. Whereas *Cdkn2a* knockdown effectively rescued the *Mph2* knockout cells from senescence, mouse Cbx7 only caused a modest delay in the onset of senescence (Figure 5.4B+C). These results are in stark contrast to the proliferation seen for the overexpression of mouse Cbx7 in wildtype MEFs and would thus suggest that the loss of *Mph2* significantly incapacitates the ability of ectopic mouse Cbx7 to repress *Ink4/Arf*.

In theory ectopic expression of Mph2 should have been rescued senescence in Mph2$^{-/-}$MEFs. In practice however, although Mph2 expression initially appeared to be modestly enhancing proliferation compared to control, this effect was not maintained (Figure 5.4B). Surprisingly, ectopic Mph2 severely impaired the proliferation of wild type MEFs (Figure 5.4C). This suggests that it might be necessary to optimise the level of expression in order to rescue the genetic defect, but this was hindered by the limited availability of cells.

5.6 Overexpression of HPH2 in HDFs

The effects of Mph2 overexpression in wild type MEFs were surprising given that ectopic expression of Cbx7 and Bmi1 has been shown to repress *Ink4a/Arf* and bypass senescence in these cells (Gil et al., 2004; Jacobs et al., 1999a). To further investigate this phenomenon, analogous experiments were performed by the overexpression of HPH2 in HDFs. To this end, mid-age Hs68 cells were infected with retrovirus containing the cDNA of *HPH2*. Upon analysis of protein and RNA levels, it was found that a modest increase (about 8-fold) in HPH2 expression leads to a strong induction of *INK4A* expression (Figure 5.5A+B). Similar findings were drawn from FDF cells, suggesting that the effect is not HDF strain specific (data not shown). The induction of *INK4A* expression correlates with strongly reduced proliferation, but the transduced cells did not fully arrest or display a senescent phenotype over time (Figure 5.5C). In the context of PRC1-mediated repression of *INK4A* locus, it was relevant to ask whether PcG occupancy of the locus was altered in the HPH2-transduced cells. ChIP analyses showed both CBX7 binding and H3K27me3 levels at the *INK4A* promoter were significantly reduced, suggesting a loss of PcG repression (Figure 5.5D).

These findings suggest that even a modest elevation of HPH2 levels can disrupt PcG-mediated repression of *INK4A*, which leads to impaired proliferation. HPH2 could possibly mediate this effect by titration of other PRC1 proteins away from *INK4A*, although this needs to be further examined. Although overexpression of PRC1 members, such as Cbx7 and Bmi1, was generally thought to produce an enhanced effect on *INK4A* repression (Gil and Peters, 2006), this might thus not universally apply to all PRC1 proteins.

5.6.1 Involvement of HPH1 and HPH3 in INK4A repression

As alluded to before, another important consideration is whether other members of the HPH family are also involved in the regulation of *INK4a*. The phenotype of the Mph1 knockout mice is similar to that of Mph2 in terms of homeotic transformations, and both proteins are proposed to have a synergistic relation in PcG-mediated repression (Isono et al., 2005a; Takihara et al., 1997). As ectopic expression of HPH1 and HPH3 proved difficult and there is an absence of reagents to study binding of the endogenous proteins to the *INK4A* locus, the main strategy employed to answer this question was shRNA-mediated knockdown.

To target HPH1 and HPH3, sets of 5 and 9 shRNAs respectively were designed and tested for efficacy by qPCR. Consequently, the best two shRNAs for each gene were used for the infection of FDF cells, alongside two shRNAs against HPH2. Depletion of both HPH1 and HPH3 resulted in an induction of *INK4A* expression both at the RNA and protein level, and to a similar degree as HPH2 (Figure 5.6). These findings therefore suggest that all HPH proteins are involved in the repression of *INK4A*. This would fit with other observations in the laboratory that multiple PRC1 complexes regulate *INK4A* and that each constituent core component is required for repression (Maertens et al., 2009)..

To further evaluate the interrelation between the different HPH proteins, combined knockdown of different pairs of HPH transcripts was attempted. For this purpose, mid-lifespan FDF cells were subjected to two rounds of infection, firstly to knockdown HPH2 and secondly to knockdown either HPH1, HPH2, HPH3 or CBX7. Following analysis of p16 expression by RT-qPCR, it was found that the effect of HPH2 knockdown could be enhanced by combination with knockdown of HPH1, HPH3 or CBX7 (Figure 5.7). Interestingly, double infection with the same HPH2 shRNA (HPH2-2) only moderately increased *INK4A* derepression. The fact that infection with empty vector control followed by HPH2-2, or vice versa had a similar effect on p16 transcript level, serves as an internal control to indicate that both plasmids were infected with similar efficiency.

These results further emphasise that repression of *INK4A* expression is achieved by the collaboration of multiple HPH proteins. With respect to the concept of an interdependent multi PRC1-complex system of PcG repression, these results may be best explained by the fact that incomplete depletion of HPH2 does not lead to full

derepression of *INK4A,* but that depletion of two proteins has an enhanced effect. The previous observations that Mph1 and Mph2 act in a synergistic manner for *Hox* gene repression in mice further supports this idea (Isono et al., 2005a). Nevertheless, these results only represent a single experiment and need to be reproduced to come to any firm conclusions.

5.7 Discussion

Following the observations in Chapter 4 that all HPH proteins interact with CBX7, but that there might be a preferential interaction with HPH2, it was unclear how the HPH proteins are involved in the demonstrated function of endogenous CBX7 or ectopic mouse Cbx7 to repress *INK4A* expression. The functional studies reported in this chapter, demonstrate that, based on shRNA-knockdown, all three HPH proteins contribute to the maintenance of *INK4A* repression. Furthermore, the inability of Cbx7 to repress *Ink4a/Arf* in the absence of *Mph2* suggests that all PH proteins are required for PcG-mediated repression. Double knockdown of two family members suggests that indeed the individual contributions to PcG-mediated repression can be targeted additively. It should however be noted that part of this data set consists of singly performed experiments and that therefore the interpretation of the data is limited and no firm conclusions can be drawn at this stage.

So far, only HPH2 has been directly associated with the *INK4A* locus, but more robust ChIP-grade antibodies under current development will allow the binding of all three proteins to be investigated. In a preliminary observation, HPH3 was also found to bind across the *INK4B-ARF-INK4A* locus with a similar distribution pattern as all other PRC1 proteins (performed by Julie Stock). It should therefore soon be possible to study the effect of HPH2 depletion on HPH3 binding to *INK4A*, or vice versa. Similarly to the findings reported in Maertens *et al.* where binding of CBX7 and CBX8 could be sequentially measured on *INK4A*, binding of both HPH1 and HPH3 can also be assessed. It will be of significant interest to determine whether the HPH proteins and their unique capacity to oligomerise could play an important part in the structural organisation and coordination of such multiple PRC1 complexes.

The observations that overexpression of HPH2 (and Mph2) functions in a dominant-negative manner on *INK4A* repression demonstrates that the basal level of HPH expression must be strictly controlled. The *Drosophila* polyhomeotic gene is

known to be tightly regulated, in part through auto-regulation by PcG complexes and by TrxG complexes, and it was proposed that the equilibrium between these states ensures an accurate level of transcription (Fauvarque et al., 1995). Furthermore, it could be that the instability of HPH2 and associated high turnover of the protein contributes to this well controlled expression level. It remains rather speculative whether consequent dissociation of CBX7 and the loss of H3K27me3 is a direct or indirect effect of increased levels of HPH2 protein. It is also important to note that Mel18 expression has been demonstrated to negatively affect cell proliferation, in part by downregulating Bmi1 expression (Akasaka et al., 1996; Guo et al., 2007a; Guo et al., 2007b; Kajiume et al., 2004), and that there have been observations in our laboratory that the overexpression of CBX8 results in an induction of *INK4A*. This suggests that this finding is not unique amongst PcG proteins and might form a basis for different functions between different family members of the PRC1 core components.

Another open question is the generalisation of the effects of HPH manipulation on other PcG target genes. Wider analysis of expression changes, for example by expression microarray, could indicate whether the effect seen for HPH2 overexpression, as well as knockdown, result in general activation of PcG target genes. However, more within the scope of this work, preliminary observations from depletion of the HPH family in the p16-defective Leiden cells suggest that the effects observed on proliferation are *INK4A*-dependent.

Figure 5.1 HPH2 is involved in the regulation of human *INK4/ARF*

A) In order to validate the knockdown of HPH2 by shRNA HPH2-1, 293T cells were co-transfected with pcDNA6 HA HPH2b and pRS HPH2-1. Whole cell lysates were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted for HA, and CDK4 was used as loading control.

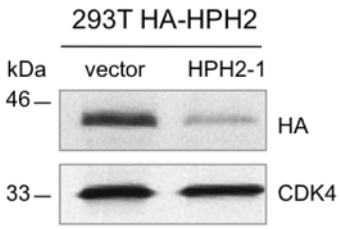
B) FDF (PD32) cells were infected with pRS HPH2-1 and pRS CBX7-1 or the empty vector control. Following 2 population doublings whole cell lysates were prepared (PD32), fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted for p16 and CDK4.

C) FDF cells were infected with pRS HPH2-1 and pRS CBX7-1 or the empty vector control. RNA was isolated 10 days after selection (when expiremental cells were at PD27 and control cells at PD35) and RT-qPCR was used to quantify the relative expression of p15, p16 and ARF, and GAPDH was used as internal standard. Error bars indicate standard deviation between unrelated biological replicates, with n=2.

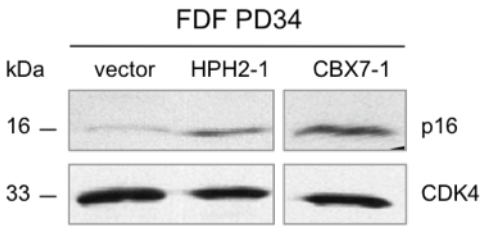
D) FDF cells as described in B) were cultured and the respective population doublings were recorded until cells transduced with HPH2-1 or CBX7-1 were both found to have arrested. This result is a representative of two separate biological replicates.

E) The FDF cells described in B) were cultured for two additional passages of the control cells, the cells were fixed and then used for staining of SA β-galactosidase activity (as described in the Materials and Methods section).

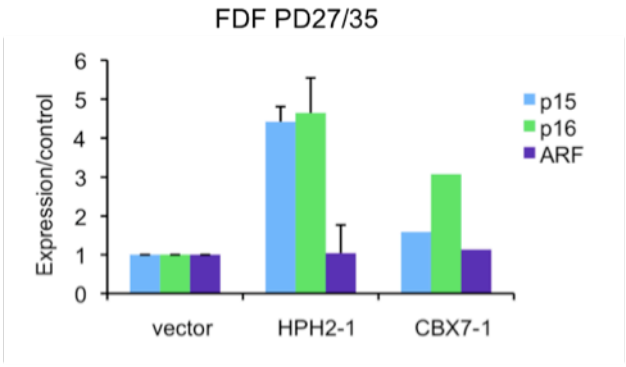
A


293T HA-HPH2
kDa
vector
HPH2-1
46
HA
33
CDK4

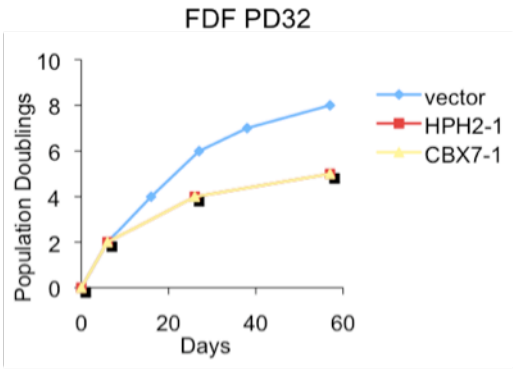
B


FDF PD34
kDa
vector
HPH2-1
CBX7-1
16
p16
33
CDK4

C


FDF PD27/35
Expression/control
0
1
2
3
4
5
6
vector
HPH2-1
CBX7-1
p15
p16
ARF

D


FDF PD32
Population Doublings
0
2
4
6
8
10
0
20
40
60
Days
vector
HPH2-1
CBX7-1

E

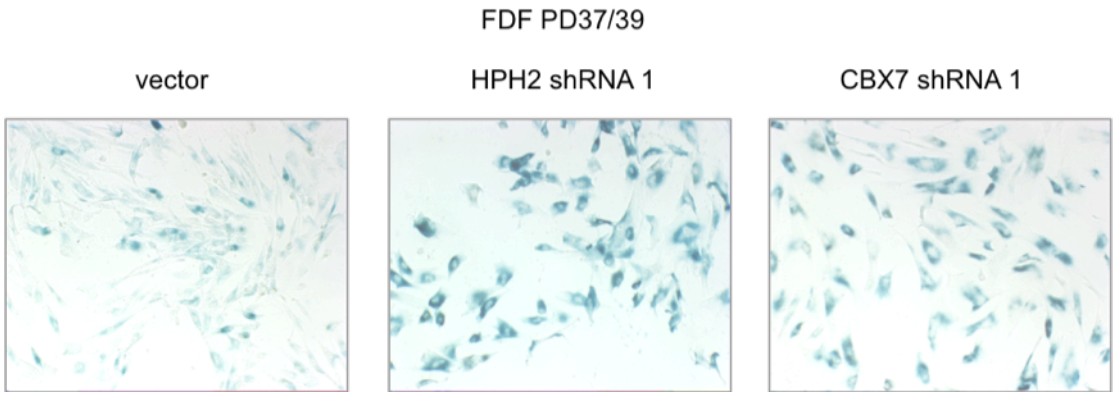

FDF PD37/39
vector
HPH2 shRNA 1
CBX7 shRNA 1

Figure 5.2 HPH2 co-localises with other PRC1 components at the *INK4B-ARF-INK4A* locus.

A) The series of primers used for interrogation of binding along the chromatin of the *INK4B-ARF-INK4A* locus is a combination of published and customly designed primers (Dietrich et al., 2007). Primer sequences are shown in section Table 5 in the Materials and Methods Chapter.

B-E) Chromatin immunoprecipitation (ChIP) was performed as described in the Materials and Methods section using antibodies for B) HPH2 (1615), C) CBX7 (Abcam), D) CBX8 (Dietrich et al., 2007) and E) BMI1 (Abcam) using chromatin of mid-life FDF cells (PD34/37). Rabbit non-specific IgG (Abcam) was included as a negative control for the rabbit polyclonal antibodies used in A, B and C, whereas, mouse anti-HA antibody was used as a negative control for the mouse monoclonal antibody used in D. Isolated DNA was interrogated by qPCR using 12 primer sets spanning across the locus as indicated in the diagram of the INK4B-ARF-INK4A locus. Error bars indicate variance between technical replicates, with n=2.

F) ChIP of H3K27me3 (Millipore) and H3 (Abcam), and qPCR analysis were performed in the same manner as described above. Due to the high signal/noise ratio of both these antibodies the rabbit unspecific IgG control is not included. The resulting enrichment of H3K27me3 was normalised to the enrichment of histone H3. A ratio of H3K27me3/H3 higher than 1 can be explained by a differential efficacy of the H3 and H3K27me3 antibodies.

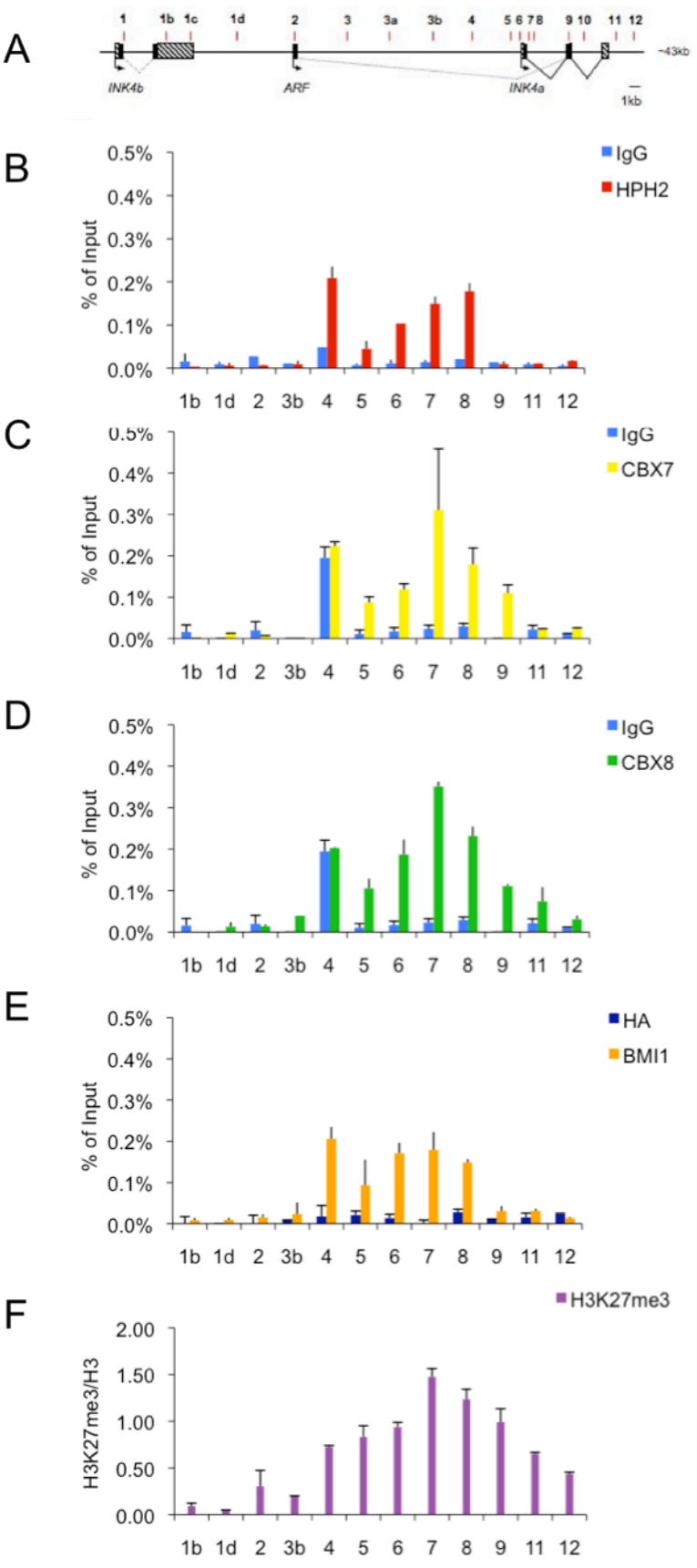

A
1 1b 1c 1d 2 3 3a 3b 4 5 6 7 8 9 10 11 12
~43kb
INK4b
ARF
INK4a
1kb
B
IgG
HPH2
% of Input
0.5%
0.4%
0.3%
0.2%
0.1%
0.0%
1b 1d 2 3b 4 5 6 7 8 9 11 12
C
IgG
CBX7
% of Input
1b 1d 2 3b 4 5 6 7 8 9 11 12
D
IgG
CBX8
% of Input
1b 1d 2 3b 4 5 6 7 8 9 11 12
E
HA
BMI1
% of Input
1b 1d 2 3b 4 5 6 7 8 9 11 12
F
H3K27me3
H3K27me3/H3
2.00
1.50
1.00
0.50
0.00
1b 1d 2 3b 4 5 6 7 8 9 11 12

Figure 5.3 HPH2 knockdown does not affect *INK4A* repression by mouse Cbx7

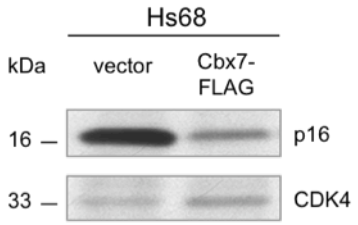
A) Mid-lifespan Hs68 (PD39) were infected with a pBABE-based virus containing mCbx7-FLAG or vector control. Cells were harvested 10 days post- selection (PD41). Whole cell lysates were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted for p16 and CDK4.

B) Cells as described in A) were used to isolate RNA for RT-qPCR interrogation of the expression levels of p16 and GAPDH (used as internal standard). Expression values plotted are normalised by the vector control.

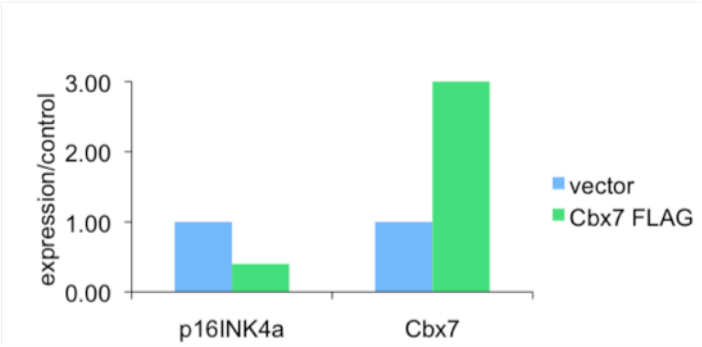
C) The cells described in A) were cultured for an additional 3 population doublings, harvested and lysed in ChIP lysis buffer (Materials and Methods). ChIP was performed using either anti-Cbx7 (Abcam) or anti-FLAG beads (Sigma). Amplifications of purified DNA are plotted as fold enrichment over the vector control.

D) HS68 described in A) were infected with shRNAs against HPH2 (HPH2-1) and CBX7 (CBX7-1). Alternatively, cells were infected with lentiviruses either empty or containing shRNA HPH2-3. Following selection, cells were harvested at PD47 and RNA was isolated. qPCR was performed with 0.02ug of cDNA of each sample to measure p16 expression. Expression of GAPDH was used as internal control. Values plotted are fold enrichment over the vector control.

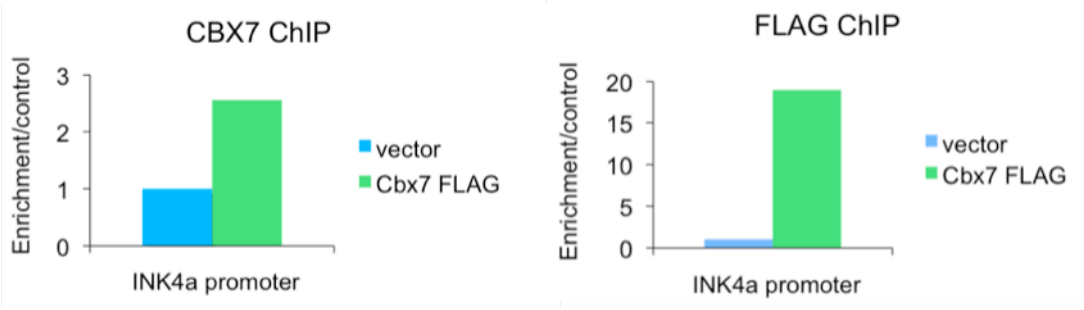
A


Hs68
kDa
vector
Cbx7-FLAG
16
p16
33
CDK4

B


expression/control
3.00
2.00
1.00
0.00
p16INK4a
Cbx7
vector
Cbx7 FLAG

C


CBX7 ChIP
Enrichment/control
3
2
1
0
INK4a promoter
vector
Cbx7 FLAG
FLAG ChIP
Enrichment/control
20
15
10
5
0
INK4a promoter
vector
Cbx7 FLAG

D

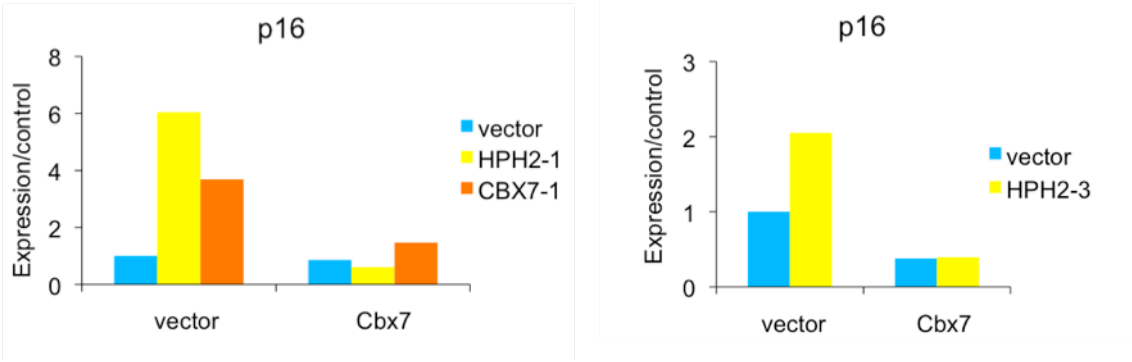

p16
Expression/control
8
6
4
2
0
vector
Cbx7
vector
HPH2-1
CBX7-1
p16
Expression/control
3
2
1
0
vector
Cbx7
vector
HPH2-3

Figure 5.4 Cbx7 repression of Ink4a is impaired in Mph2-/- MEFs

A) MEFs were infected with a pMARX retrovirus encoding Cbx7-FLAG, Mph2 or vector control, or the pMLP retrovirus containing an shRNA against *Cdkn2a*. After 7 days, cells were harvested and lysed, extracts were fractionated by SDS PAGE, transferred to nitrocellulose and immunoblotted with anti-p16 (Santa Cruz) or anti-Cbx7 (Abcam).

B) Wildtype and Mph2-/- MEFs were infected with pMARX encoding Mph2 (cDNA), Cbx7-FLAG or vector control, or pMLP *Cdkn2a* shRNA. After selection, and at the start of population doubling measurements, cells were photographed at 50x magnification.

C) Cells described in B, were cultured and their population doublings were recorded over time.

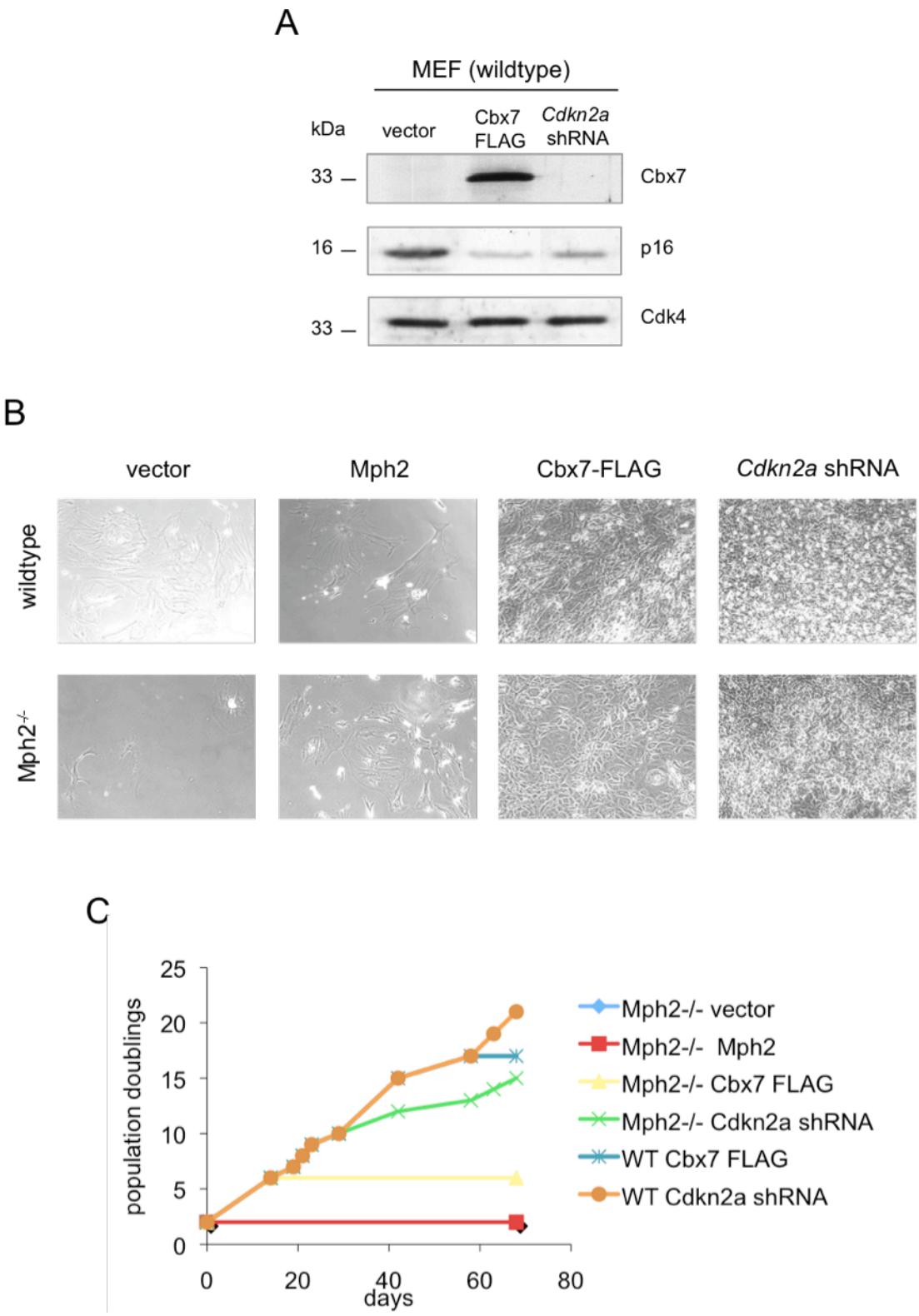

A
MEF (wildtype)
kDa
vector
Cbx7 FLAG
Cdkn2a shRNA
33
Cbx7
16
p16
33
Cdk4
B
vector
Mph2
Cbx7-FLAG
Cdkn2a shRNA
wildtype
Mph2-/-
C
population doublings
25
20
15
10
5
0
0
20
40
60
80
days
Mph2-/- vector
Mph2-/- Mph2
Mph2-/- Cbx7 FLAG
Mph2-/- Cdkn2a shRNA
WT Cbx7 FLAG
WT Cdkn2a shRNA

Figure 5.5 HPH2 overexpression activates *INK4A* by depletion of PcG-mediated repression

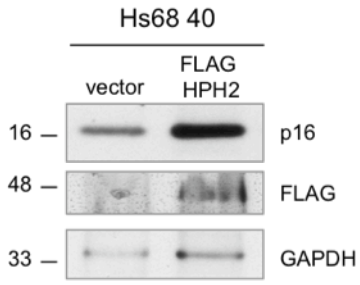
A) Mid-life Hs68 cells were infected with virus containing pBABE FLAG HPH2 or vector control. 6 days following infection, cells were harvested and lysed. Whole cell lysates were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted fpr p16, FLAG or GAPDH (Sigma).

B) RNA was extracted from the same batch of cells as in A) and RT-qPCR was used to quantitate the expression of p16, HPH2 and GAPDH (internal standard). Error bars indicate standard deviation between independent biological replicates, with n=2.

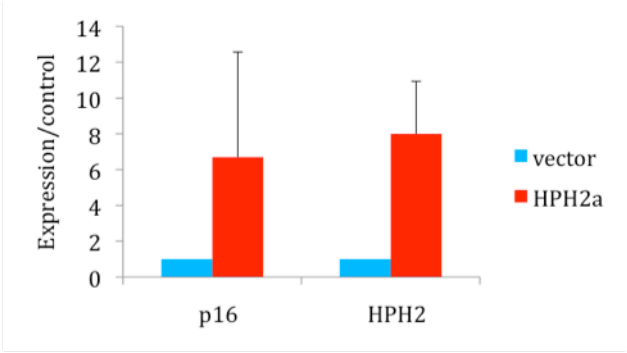
C) Hs68 cells as described in A) were cultured and the respective population doublings were recorded and plotted over time. This is a representative of results from two distinct cell types.

D) The Hs68 cells described in A) were used for ChIP analysis by precipitating CBX7 and H3K27 and sampling purified DNA with primers to the *INK4A* promoter (PS6). This figure represents the combined results of two independent biological replicate experiments.

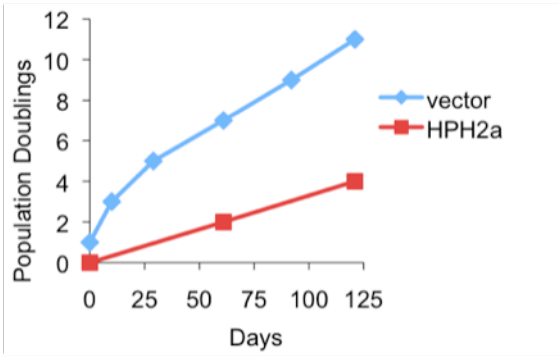
A


Hs68 40
vector
FLAG HPH2
16
48
33
p16
FLAG
GAPDH

B


Expression/control
14
12
10
8
6
4
2
0
p16
HPH2
vector
HPH2a

C


Population Doublings
12
10
8
6
4
2
0
0
25
50
75
100
125
Days
vector
HPH2a

D

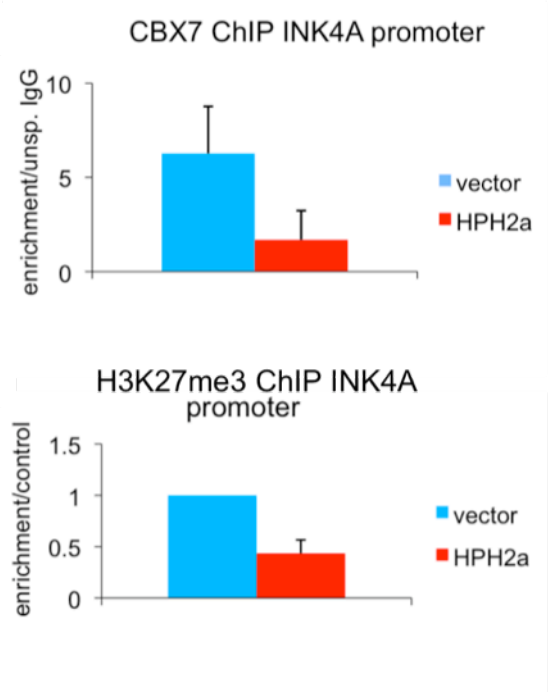

CBX7 ChIP INK4A promoter
enrichment/unsp. IgG
10
5
0
vector
HPH2a
H3K27me3 ChIP INK4A promoter
enrichment/control
1.5
1
0.5
0
vector
HPH2a

Figure 5.6 All HPHs are involved in *INK4A* regulation

A) Hs68 cells (PD45) were infected with retroviruses encoding shRNAs against HPH1 (H1-1 and H1-2), HPH2 (H2-1 and H2-2) and HPH3 (H3-1 and H3-2) and were analysed by qPCR (as described section 2.5.2) for the expression of HPH1, HPH2 and HPH3 respectively, and all for p16 and GAPDH (internal standard).

B) Whole cell lysates of Hs68 cells as described in A) were fractionated by SDS-PAGE, transferred to nitrocellulose and immunoblotted for HPH1 (Abnova), HPH2a (Abcam) and HPH3 (Marc F. Hansen) for each respective knockdown, and p16 and CDK4.

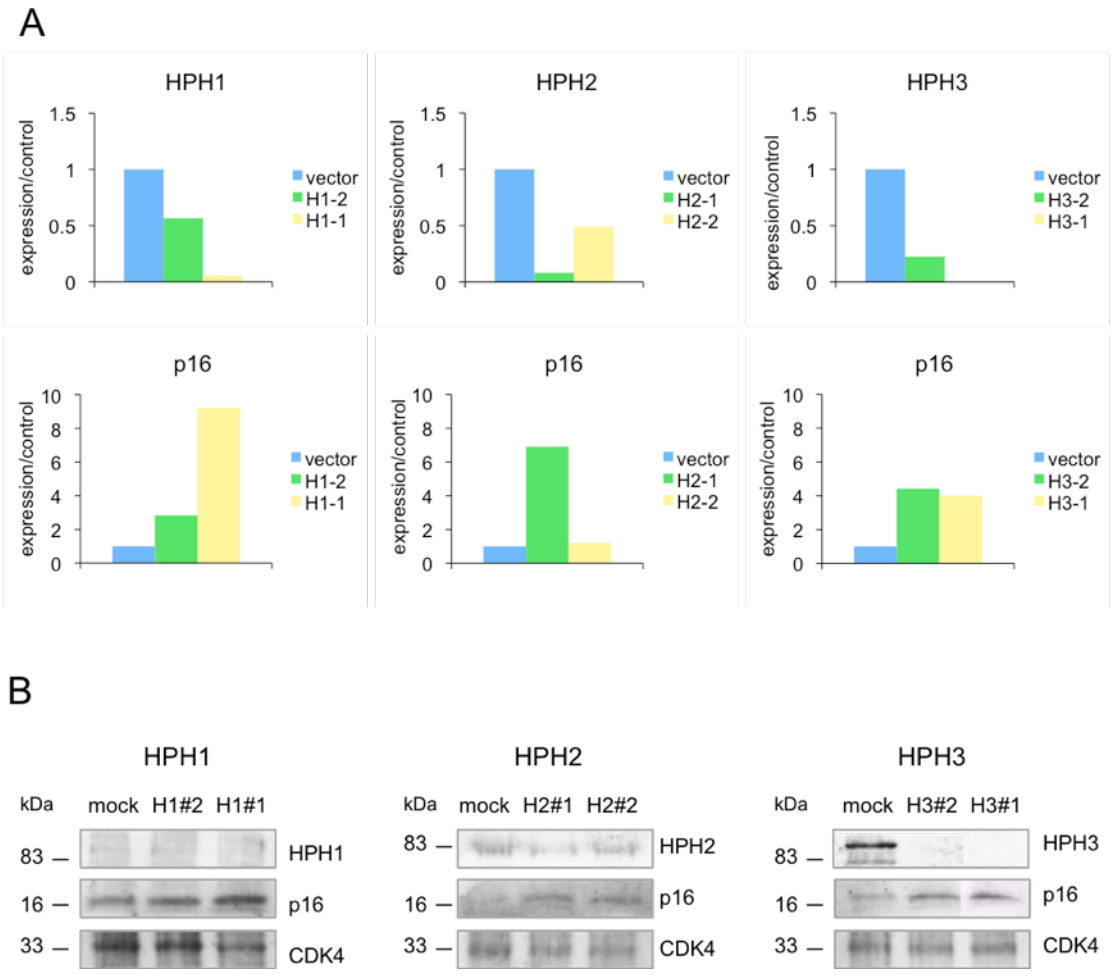
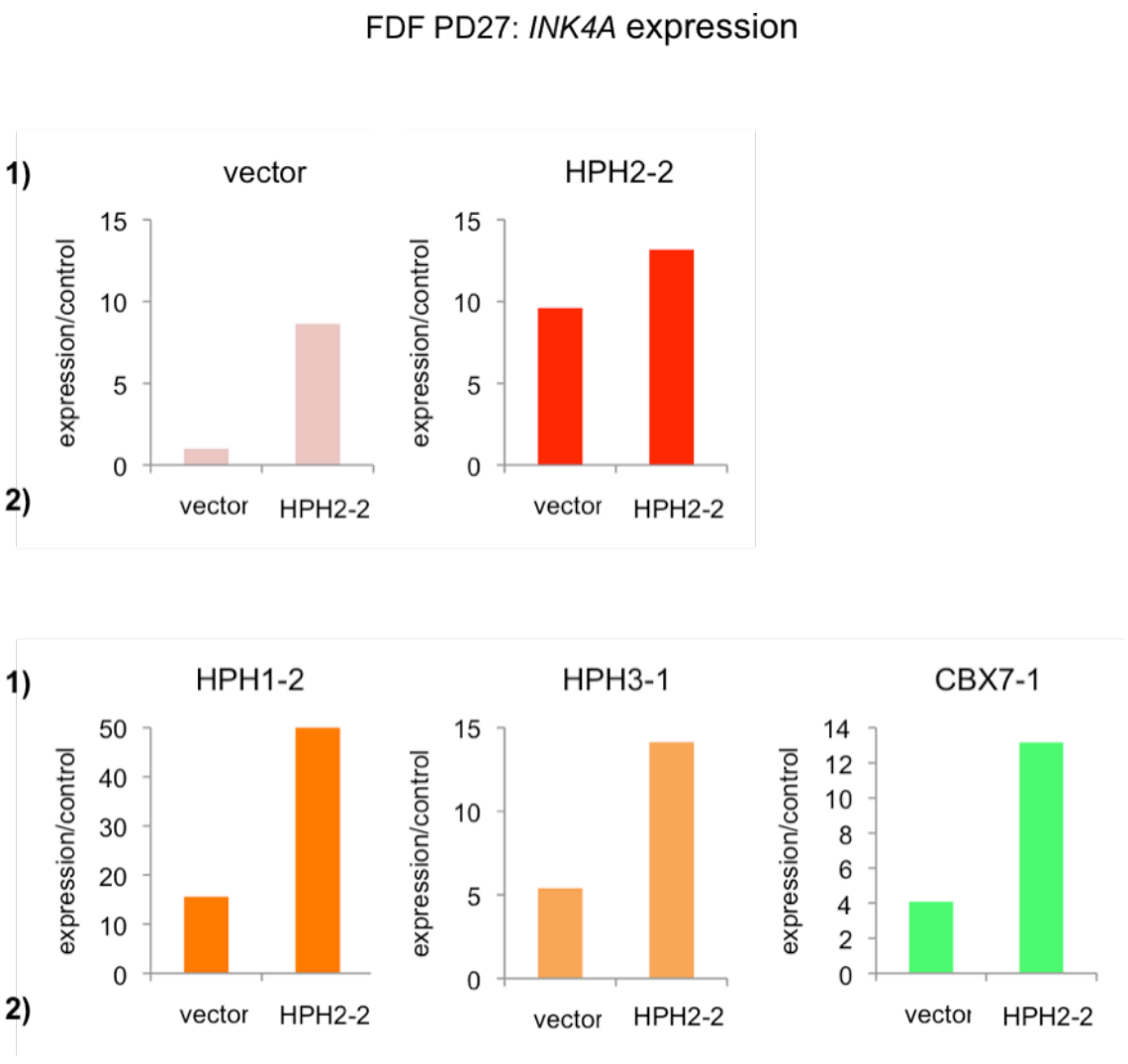


Figure 5.7. Depletion of two HPH proteins suggests synergy in repression

Mid-age FDF cells were subjected to two rounds of infection and selection with 1) pRS hygromycin empty or containing HPH2 shRNA-3 (HPH2-2), and 2) pRS puromycin empty or containing HPH2-2, HPH1-2, HPH3-1 or CBX7-1. Following the second selection cells were harvested for RNA isolation. 1ug of RNA was used for RT-qPCR for p16 and GAPDH (internal standard) expression. Resulting expression levels were normalised by the double empty vector infected control.



6 The Functional Role of the Human Polyhomeotic Proteins in the Regulation of *INK4A*

6.1 Introduction

As previously discussed, the HPH proteins have a number of properties that could carry an important role in the mechanism of PcG-mediated repression. These properties include RNA-binding, oligomerisation and the interaction with MAPK and DDR protein kinases, which altogether suggest both regulatory and structural roles for the HPH family. However, the conserved functional domains are clustered in the C-terminal region of the HPH proteins and no functional motifs have directly been associated with the extensive amino terminal region. Understanding how these C-terminal regions confer functionality can lead to new insights into the functionality of the different HPH proteins, and also could be instrumental for differentiating between the roles of the HPH2a and HPH2b isoforms.

6.2 Identification of novel Interactions of HPH2

As discussed in Chapter 4, the overlap between HPH2a and HPH2b has made it difficult to assess the individual functions of each, as there is no possibility of a specific antibody, nor can the HPH2b isoform be specifically amplified by PCR. As an alternative approach, it was decided to try and purify both proteins and compare the interacting proteins identified. To this end, the classical TAP tag (Rigaut et al., 1999) was cloned on the C-terminus of human HPH2b. This plasmid and TAP-tagged mouse Cbx7 (Figure 3.1A) were then transiently transfected in 293T cells and used for a pilot scale purification, carried out using IgG sepharose beads followed by TEV cleavage and a second round of purification on calmodulin beads. Figure 6.2A shows that no interacting bands were observed on a silver stained gel for HPH2b, whereas purification of Cbx7 was found to produce its characteristic pattern of bands (Figure 4.1A). This was observed for small-scale purifications from both transiently transfected 293T and DU-145 cells. One of the possible explanations was that the C-terminal tag would interfere with the functional properties of HPH2b.

Based on this suspicion, another cloning strategy was employed as to create the TAP tag on the amino terminus instead, and constructs were made for both HPH2a and HPH2b. Small-scale purifications following transient transfection with these constructs, were separated by SDS-PAGE and silver stained. Following staining a clear set of bands was visible for both isoforms, which are both indicated by an arrow (Figure 6.2B). These include shared bands, but also clearly distinct ones, which were not present in the TAP only control. This therefore suggested that this method may be employed to characterize the unique properties of these proteins.

As HPH2a is likely to boost the most interactions due to its significantly larger size, a large scale purification of this protein was pursued. Several attempts to perform large scale purifications from transiently transfected cells only produced limited yield and no detectable bands following Colloidal Coomassie staining, which has been the general criterion in the laboratory for pursuing identification by mass spectometry. As scale appeared to be the limiting factor, which may be attributed to low stability of HPH2a throughout the purification procedure, it was decided to use a large-scale clonal culture. To this end, 293T cells were stably transfected with HPH2a with the N-terminal TAP tag. 42 clones were derived and tested for individual expression levels. The best expressing clone was consequently used for expansion (performed by LRI Cell Services) and 50 ml of cells were used to achieve sufficient purification yield for downstream analysis. The purified proteins were then separated by SDS-PAGE, stained by either SYPRO Ruby (purification 1) or Coomassie Colloidal Brilliant Blue (purification 2) (Figure 6.3). Protein gel bands were manually excised, and samples were consequently trypsinised and analysed by LC-MS/MS (performed by the LRI Protein Analysis and Proteomics laboratory).

The two purifications yielded 169 and 68 identified proteins, respectively, and had a reasonable amount of overlap (39 proteins) (

Table 9 and Table 10). Although the same amount of starting material was used for both purifications, the difference in the number of identifications and the number of peptides for mutual identifications suggests that the second purifications was overall less efficient or that the analysis performed was less sensitive. The fact that the purified proteins were visualised by different stains does not allow for direct comparison of yield. The large total number of different proteins identified is a reason for concern, especially since no TAP alone clone was available to use as a negative control for this scale of purifications, and therefore no background identifications can be reliably

discounted. However, the reasonable number of reproducible indentifications, especially of other PcG proteins, is reassuring.

The resulting list of proteins co-purifying with HPH2a includes both PRC1 and non-Polycomb proteins. It is confirmatory to find that at least one representative of each PRC1 core component was found to co-purify, which suggests that intact PRC1 complexes were purified. In addition to RING1B, RING1A was also found in the first, possibly more sensitive purification, with 2 peptides in contrast to 13 for RING1B. The possible co-purification of both RING1 and RING2 as well as both CBX2 and CBX8 suggests that HPH2 is indeed involved in multiple PRC1 complexes. This would fit with the findings in Chapter 4 that suggest that HPH2a has no specificity for any particular member of the CBX family. In this context, it is important to consider that the composition of PRC1 is likely to be influenced by the expression levels of individual proteins in 293T cells. CBX7 is only expressed at low level in 293T cells and therefore may not have been identified (Bernard et al., 2005). The same may possibly apply to CBX4 and CBX6.

The non-PRC1 proteins identified represent a diverse range of proteins. Some are likely contaminents, such as keratin and immunoglobulin chains. Others are known to frequently co-purify unspecifically, including a large number of ribosomal proteins, heat shock proteins (HSPs) and histone proteins. The purifications also yielded many proteins which fall within the 'sepharose proteome', which refers to proteins which commonly interact directly with the sepharose beads used in many purification protocols, includes numerous RNA binding proteins, such as heterogeneous nuclear ribonucleoproteins (hnRNPs), ATP-dependent RNA helicases and splicing factors (Trinkle-Mulcahy et al., 2008). The identified proteins suspected to bind sepharose according to this resource are highlighted in light yellow in Table 9 and Table 10. However, as HPH proteins are expected to bind RNA via the conserved FCS domain, it would be premature to fully dismiss these proteins as contaminents (Zhang et al., 2004). An interesting example of this would be the identification of RNA helicase A (RHA), which has been shown to specifically bind to the promoter of *INK4A* and mediate activation of the locus (Myohanen and Baylin, 2001). This would thus suggest that there is possible co-localisation of RHA and HPH2 at the *INK4A*, an interesting hypothesis that remains to be tested. Furthermore, the following proteins were reproducibly identified to bind to HPH2 and are not described to be general contaminents;

Nucleolar protein 6 and UT14A are both involved in the processing of pre-rRNA primary transcripts (Rohozinski and Bishop, 2004; Utama et al., 2002). RNA

binding motif protein 14 (RBM14/CoAA) is involved in the regulation of alternative splicing and is overexpressed in a subset of human tumours (Sui et al., 2007). Interestingly, RBM14 is also reported to be a substrate of ATM/ATR phosphorylation, in conjunction with HPH2 (Matsuoka et al., 2007). Splicing factor 3B subunit 4 (SF3B4) is also involved in splicing as SF3B is part of the U2 small nuclear ribonucleoprotein particle, a subcomponent of the splicosome. SF3B subunit 2 was also identified in the second purification (4 peptides). The co-purification of these subunits is particularly interesting since subunit 1 of Sf3b has been reported to be essential for mammalian PcG repression of *Hox* genes and to interact with a number of PcG proteins, including mouse Mph1, Mel18 and Ring2 (Isono et al., 2005b). This report also found the interaction of Sf3b4 with all PRC1 proteins tested (Mel18 and Ring2). Altogether this strongly suggests an interaction between the SF3B and PRC1 complexes.

Furthermore, the YLP-motif containing protein 1 (YLMP1/ZAP3) is a putative nucleoside kinase that targets protein phosphatase-1 (PP1) and interacts with a number of RNA binding proteins, including a member of the hnRNP family (Ulke-Lemee et al., 2007). Interestingly, PP1 was shown to dephosphorylate SF3B1 through the action of the nuclear inhibitor of PP1 (NIPP1) protein, which leads to the activation of splicing (Tanuma et al., 2008). Strikingly, NIPP1 has been reported to interact directly with EZH2 and SUZ12 and to be regulated by the associated methylation activity of these proteins (Roy et al., 2007).

The co-purification of histone H1.2 is also intriguing. Although histone proteins are regarded as possible contaminents, histone H1.2 is involved in mediating higher-order chromatin structures and depletion of this specific H1 variant results in a G1-phase arrest, correlated with expression changes of cell cycle genes, including downregulation of *CDKN2A* and *CDKN2B* expression (Sancho et al., 2008). Furthermore, loss of histone H1 has an important role in cellular senescence by allowing the formation of SAHFs. These findings suggest that H1.2 could have a role in PcG repression of *INK4A* and warrants further investigation.

The Thyroid hormone receptor-associated protein 3 (THRAP3/TRAP150) is a component of the human mediator coactivator complex that interacts with transcriptional activator TFIID (Johnson et al., 2002). This is particularly relevant, as *Drosophila* PRC1 proteins, including polyhomeotic, were found to robustly associate with TFIID (Saurin et al., 2001; Wang and Brock, 2003). Furthermore, amongst the three best individual hits from the second purification were Polymerase delta-interacting protein 3 (3 peptides) and Transcription elongation factor A protein-like 4 (2 peptides).

Altogether this would suggest a possible link between HPH2 and the transcription machinery.

7,8-dihydro-8-oxoguanine triphosphatase (8ODP/NUDT1/MTH1) is an enzyme that hydrolyzes oxidized purine nucleoside triphosphates, to monophosphates, thereby preventing misincorporation into DNA/RNA during replication and transcription (Mo et al., 1992; Nakabeppu et al., 2006). It is overexpressed in a number of cancers, potentially as a result of increased reactive oxygen species (ROS) (Kennedy et al., 2003). Interestingly, MTH1 depletion was reported to lead to premature-senescence with elevation of $p21^{cip1}$, p53 and $p16^{INK4A}$ (Rai et al., 2009). Although the identification of both these proteins are not intuitive based on conventional PcG literature, recent studies have reported mitochondria- and ROS-related genes to be regulated by PcG and that *Bmi1*$^{-/-}$ mice have impaired mitochondrial function and elevated levels of intracellular ROS, suggesting that PRC1 could be involved in the regulation of mitochondrial function and ROS generation (Bracken et al., 2006; Liu et al., 2009).

Although heat shock proteins (HSPs) are commonly regarded as contaminents in protein purification (Trinkle-Mulcahy et al., 2008), the identification of several HSPs in the HPH2 purifications may be connected to a number reported findings. Firstly, HSP70, was previously identified as a component of PRC1 (Levine et al., 2002; Saurin et al., 2001). Furthermore, PcG might also be involved in the regulation of HSP70, as in *Drosophila* Pho was shown to be required for transcriptional repression of Hsp70 following heat shock (Beisel et al., 2007). In addition to HSP70, heat shock cognate 4 (Hsc4) has shown a genetic interaction with PcG genes and Hsc4 stably associates with polyhomeotic in *Drosophila* (Mollaaghababa et al., 2001; Wang and Brock, 2003). Also known interacting proteins of polyhomeotic are associated with HSPs; MK2 was found to phosphorylate stress-dependent small HSPs (Vertii et al, 2006).

In summary, the purification of HPH2 has provided insight into possible preferential PRC1 interactions and a number of putative novel interactions. There is a prominence of proteins involved in RNA processing, which could directly be connected to the RNA binding capacity of HPH2. Furthermore, the possible connection of HPH2 to linker histones, basal transcription factors and mitochondrial function would suggest novel additional roles for the protein.

HPH2 interactions with RNA

As discussed in Chapter 4, the RNA binding domain of HPH proteins is highly conserved and is thought to mediate the interaction between the HPH and CBX families. Several studies have implicated non-coding RNA (ncRNA) in the regulation of the PcG target genes (Khalil et al., 2009; Pandey et al., 2008; Rinn et al., 2007; Zhao et al., 2008). It was therefore investigated whether ncRNA is involved in the repression of *INK4A* and whether this would involve the RNA binding properties of HPH2.

6.2.1 Profiling of antisense transcription along the *INK4B-ARF-INK4A* locus

Several antisense ncRNAs have been identified to originate from or read through the *INK4B-ARF-INK4A* locus (Gonzalez et al., 2008; Guttman et al., 2009; Yu et al., 2008), but little is known about their likely structure or regulation. Since a comprehensive assessment of *INK4B-ARF-INK4A* non-coding transcripts was lacking, a low-resolution profile was made of sense and antisense transcription of the *INK4B-ARF-INK4A* locus using a 'sense/antisense specific RT-qPCR' method (Figure 6.4). The previously mentioned *INK4B-ARF-INK4A* genomic primer sets, providing coverage of the locus at low resolution, were employed. By performing reverse transcription (RT) using the forward primer of one of these primer pairs, any RNA transcript antisense with respect to the locus is specifically reverse transcribed from total RNA. Conversely, using the reverse primer of a pair in the RT reaction, any transcript sense with respect to the locus is reverse transcribed. The resulting cDNA was then purified on a PCR purification column (Qiagen) to ensure no carry-over of RNA and RT reaction components. In the next step the cDNA is amplified by qPCR using both the forward and the reverse primer (one of which was originally used for the RT reaction). Using qPCR to amplify the series of RT reactions corresponding to the different target regions along the *INK4/ARF* locus, a relative expression pattern across the locus is produced. Differences in primer amplification efficiency were accounted for by normalisation to the signal from amplification of genomic DNA.

This procedure was performed to map sense- and antisense transcripts of the *INK4B-ARF-INK4A* locus with total RNA from Hs68 and IMR90 (female fetal lung)

HDFs. Both cell types produced a very similar profile. High levels of sense transcription could be found at the promoter of *INK4A* and smaller peaks were observed along the rest of the gene. A relative abundance of antisense transcription relative to the promoter and exon 1 region of *INK4A* was found (Figure 6.5). Interestingly, these peaks overlap with the enrichment of PcG proteins on the locus (as demonstrated in Figure 5.2).

6.2.2 HPH2 binds transcripts antisense to the *INK4A* promoter

As the antisense transcription and binding of the PRC1 proteins appear to co-localise, the possible RNA binding to PRC1 was studied. For this purpose, RNA chromatin immunoprecipitation (RNA ChIP) was employed. Since the HPH and CBX proteins are the candidate RNA binding proteins in the PRC1 complex and antibodies for CBX7 and HPH2 were available, these 2 proteins were specifically evaluated. Briefly, the RNA ChIP technique employs the same principles as conventional ChIP, but focus is on ensuring the stability and recovery of RNA molecules rather than DNA molecules throughout the procedure. Since the protocol involves a formaldehyde cross-linking step, the possible contamination of genomic DNA was minimised by two independent RNase-free DNase treatments. In addition, RT control reactions, lacking the RT enzyme, were performed for each sample, and consistently did not render any detectable amplifications in qPCR analysis. Following precipitation of the RNA, it was investigated whether any antisense RNA transcripts corresponding to the *INK4A* promoter region (primer set 6) could be amplified. This was based on the overlap of HPH2 or CBX7 binding and the high level of transcription at this site. In contrast to CBX7, HPH2 was found to reproducibly co-purify RNA transcripts antisense to the *INK4A* promoter (Figure 6.6). Due to the low yield of RNA obtained from this procedure, the binding at other locations could not be examined.

Following the observation that HPH2 interacts with RNA antisense to the *INK4A* promoter, it was questioned whether mutation of the RNA binding capacity of HPH2 affects the induction of *INK4A* following HPH2 overexpression. To this end, mutants of the conserved cysteine residues of the FCS domain, cysteine 1 and 2 (C1C2) or all 4 cysteines (C1-4) were constructed as previously characterised (Zhang et al., 2004) (kindly provided by Marc Rodriguez-Niedenführ). These mutants were then overexpressed in Hs68 alongside wildtype HPH2. Evaluation of *INK4A* induction by

qPCR demonstrates that the induction of *INK4A* expression was not found to be affected by the loss of RNA binding capacity of HPH2 (Figure 6.7). Furthermore, no differential effect on proliferation was observed for these mutants in comparison to wildtype (empirical observations). At this stage, the experiment was only conducted once and therefore no conclusions can be drawn based on these initial findings. However, this data would suggest that RNA binding is not required for the, as yet uncharacterised, mechanism of *INK4A* induction by ectopic expression HPH2. Therefore, the question remains open on what the mechanistic importance of the interaction between HPH2 and *INK4A* antisense transcripts is.

6.3 Discussion

As the polyhomeotic component remains the most elusive of the PRC1 complex, it was aimed that investigation of the properties of HPH2, preferably in the direct context of *INK4A* repression, would shed more light on HPH functions in regulating a particular target gene.

Purification of HPH2 from 293T cells has provided a general list of putative interactions that might eventually contribute to the mechanism of *INK4A* repression. This set of interacting proteins strongly suggests a connection to the RNA binding function of the FCS domain of HPH2 and it will be of much interest to further understand any underlying mechanisms. In addition, the putative interactions between HPH2 and the SF3B subunits, as well as YLPM1/ZAP3, provide a hypothesis on how HPH2 and other PcG proteins may be involved in the regulation of splicing. The SF3B complex is an important component of the splicosome. SF3B subunit 4 reproducibly co-purified with HPH2, whereas SF3B subunit 2 co-purified with HPH2 once. In addition, the interaction between PRC1 proteins and SF3B subunit 1 has previously been reported (Isono et al., 2005b). ZAP3, a co-factor of PP1, also reproducibly co-purified with HPH2 (Ulke-Lemee et al., 2007). ZAP3 has been reported to interact with hnRNP G, which co-purified once with HPH2 (Ulke-Lemee et al., 2007). Furthermore, NIPP1 has been reported to interact with PRC2 (Roy et al., 2007), whereas the PRC2-associated protein NURF55 was co-purified once with HPH2 as well. Altogether these published and novel findings can be structured as displayed in Figure 6.1.

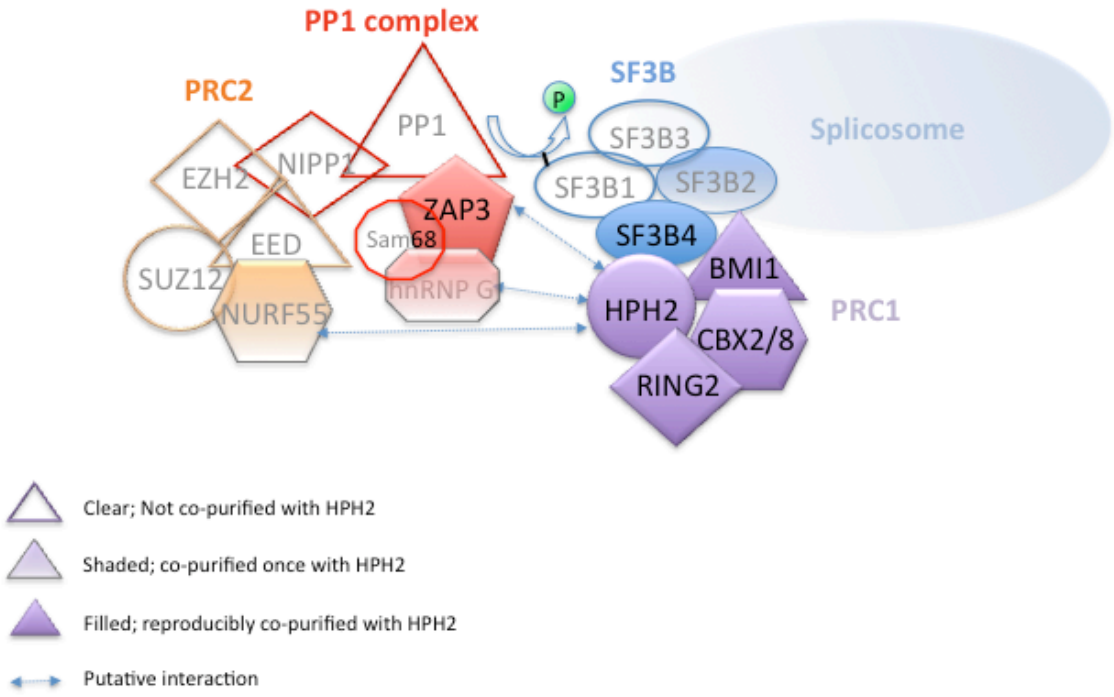


Figure 6.1 Putative hypothesis for the regulation of splicing in the context of PcG complexes

NIPP1-mediated dephosphorylation by PP1 of SF3B subunit 1 (Tanuma et al., 2008) may occur in the context of PRC1 and PRC2 proteins based on previously reported findings and the data presented in this work. An unfilled shape represents a protein not co-purified with HPH2, a shape with shaded fill indicates a protein co-purified with HPH2 once, and a filled shape indicates a protein co-purified twice with HPH2.

Although the novel interactions identified by tandem affinity purification have interesting implications, a number of known HPH2-interacting proteins were not identified through this approach. Examples of these include MK2/3, SP1 and ATM/ATR. Nevertheless, these interactions have either been found *in vitro* and/or are likely to be of a transient nature and may therefore not have sustained throughout the purification procedure. Additionally, it is important to consider the artefacts that may have arisen from the supraphysiological levels of HPH2. Stably transfected 293T cells with a more modest level of expression would have been preferred in this regard, but would have further compromised the yield of the purification.

Furthermore, investigation of RNA binding and transcription sense and antisense to the *INK4B-ARF-INK4B* locus has led to several interesting observations. First of all it was found that there is clearly a strong level of transcripts antisense to the

INK4A promoter and exon 1 region, which is not affected at senescence when the locus is strongly expressed. It is furthermore striking that these transcripts correspond to the peaks of PRC1 protein binding, which would suggest that the two events might be linked. The fact that HPH2, but not CBX7, was found to co-immunoprecipitate transcripts antisense to the *INK4A* promoter is therefore compelling and needs to be further investigated in order to define a possible mechanism.

Given the time restraint to complete this work, only the surface of HPH functionality in the context of *INK4A* regulation has been scratched. It will be important to dissect the role that the identified interactions have in the regulation of *INK4A* and possibly extend this to the known related mechanisms, especially with respect to RNA processing. As it is becoming increasingly evident that RNA is linked to PcG repression, it will be important to understand the nature of antisense (and sense) transcripts that match to *INK4A* and understand their role, origin and regulation. As it appears that transcripts of RNA detected are of a relatively short length, it is difficult to target them by RNA interference without disturbing the coding transcripts originating from the locus. Therefore, the aim should be to characterise these transcripts by full transcriptome sequencing and consequently be able to ectopically express the transcripts to determine the effect on *INK4A* repression.

Figure 6.2 Evaluation of HPH2 Tandem Affinity Purification

A) DU-145 cells were transiently transfected with pcDNA3 encoding Cbx7-TAP, HPH2b-TAP or TAP alone. The cells were lysed in NP40 lysis buffer and a small scale TAP purification was carried out as described in section 2.4.6. The purified proteins samples were fractionated by SDS-PAGE and the gel was stained using SilverQuest Silver Staining kit (Invitrogen). Similar results were obtained using 293T cells (data not shown).

B) 293T were transiently transfected with pcDNA6 encoding TAP-HPH2a, TAP-HPH2b or TAP alone. Purification was performed as described in A).

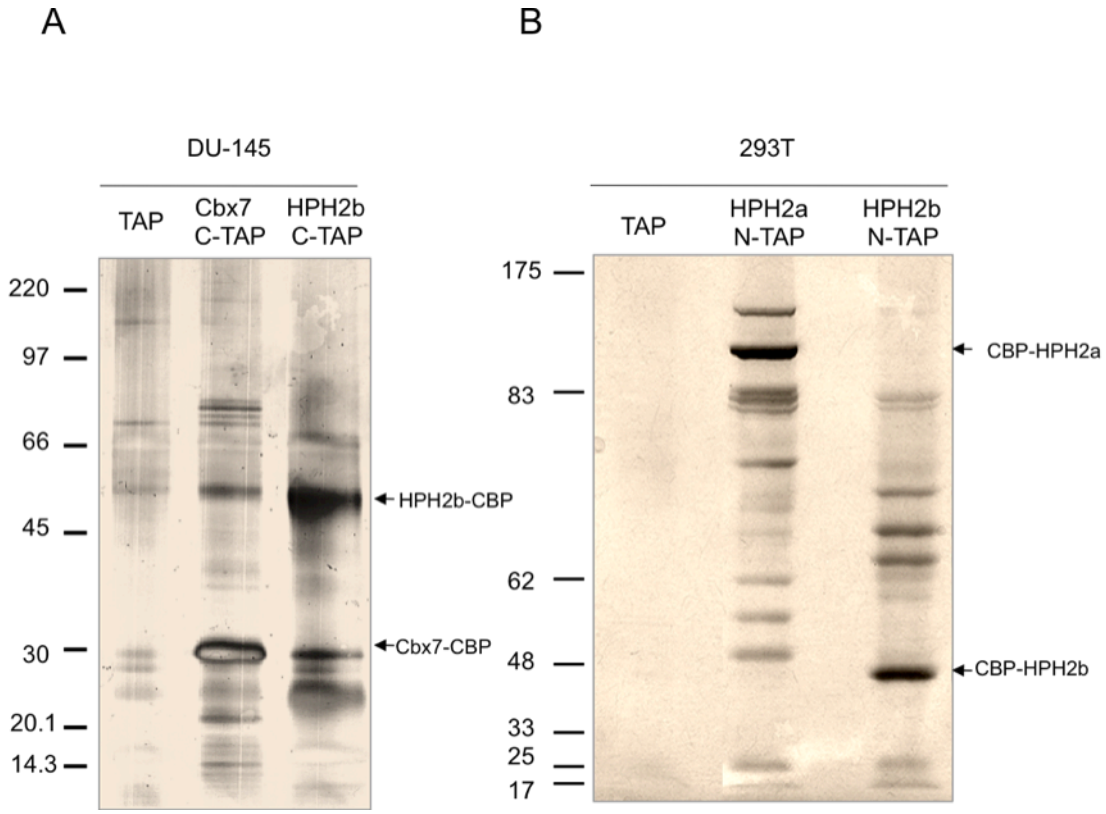
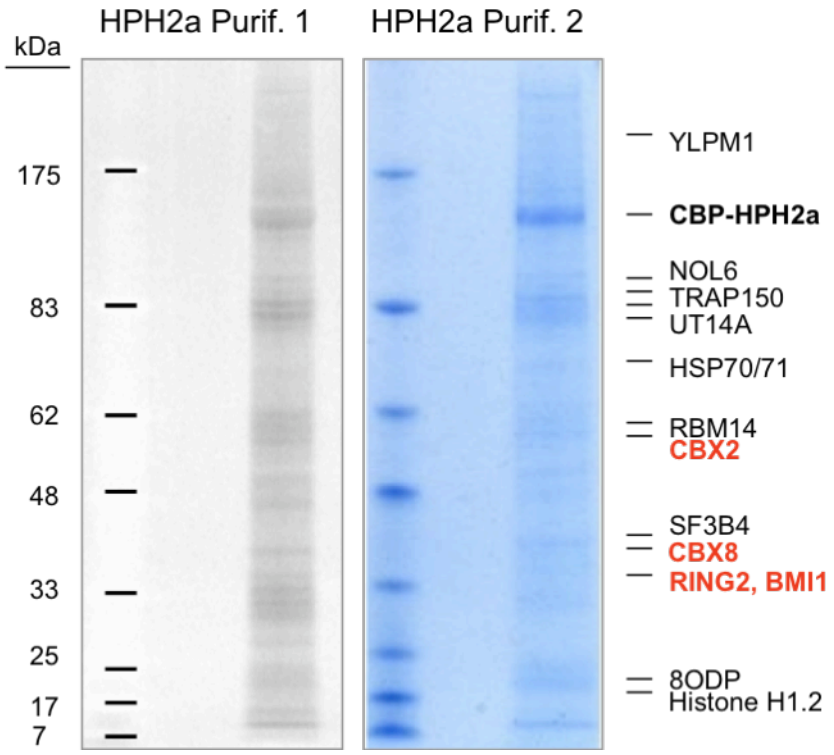


Figure 6.3. Identifcation of HPH2a interacting proteins

Large scale tandem affinity purification was carried out two consecutive times using 50ml of 293T cells stably expressing HPH2a with an N-terminal TAP-tag. The resulting proteins were fractionated by SDS PAGE on a Novex® 4-20% Tris-Glycine gel (Invitrogen) and stained with SYPRO ruby (first purification) or Coomassie Briliiant Blue (Sigma) (second purification). Visible bands were excised for analysis by mass spectrometry. Reproducibly identified proteins not suspected to be likely contaminants are indicated by name, with PRC1 proteins in red and CBP-HPH2a in **bold**.



Identified Proteins	MW	P1	Prob1	P2	Prob2
Polyhomeotic-like protein 2 - Homo sapiens (Human)	91 kDa	14	100%	6	100%
E3 ubiquitin-protein ligase RING2 - Homo sapiens (Human)	38 kDa	13	100%	1	95%
Histone H1.2 - Homo sapiens (Human)	21 kDa	6	100%	2	91%
Chromobox protein homolog 8 - Homo sapiens (Human)	43 kDa	5	100%	5	100%
Nucleolar protein 6 - Homo sapiens (Human)	128 kDa	5	100%	5	100%
YLP motif-containing protein 1 - Homo sapiens (Human)	220 kDa	4	100%	1	75%
RNA-binding protein 14 - Homo sapiens (Human)	69 kDa	3	85%	3	99%
Thyroid hormone receptor-associated protein 3 - Homo sapiens (Human)	109 kDa	3	80%	2	98%
Chromobox protein homolog 2 - Homo sapiens (Human)	56 kDa	2	92%	3	100%
Polycomb complex protein BMI-1 - Homo sapiens (Human)	37 kDa	2	100%	1	94%
U3 small nucleolar RNA-associated protein 14 homolog A - Homo sapiens (Human)	88 kDa	1	53%	3	100%
7,8-dihydro-8-oxoguanine triphosphatase - Homo sapiens (Human)	23 kDa	1	87%	1	90%
Splicing factor 3B subunit 4 - Homo sapiens (Human)	44 kDa	1	84%	1	100%
Heat shock 70 kDa protein 1 - Homo sapiens (Human)	70 kDa	12	100%	3	100%
Heterogeneous nuclear ribonucleoprotein M - Homo sapiens (Human)	78 kDa	10	100%	2	100%
Heat shock cognate 71 kDa protein - Homo sapiens (Human)	71 kDa	9	100%	7	100%
Heterogeneous nuclear ribonucleoprotein U - Homo sapiens (Human)	91 kDa	9	100%	5	100%
Matrin-3 - Homo sapiens (Human)	95 kDa	6	89%	2	95%
Probable ATP-dependent RNA helicase DDX41 - Homo sapiens (Human)	70 kDa	5	100%	2	94%
Stomatin-like protein 2 - Homo sapiens (Human)	39 kDa	4	100%	1	95%
60S ribosomal protein L23a - Homo sapiens (Human)	18 kDa	4	100%	1	91%
Electron transfer flavoprotein subunit beta - Homo sapiens (Human)	28 kDa	2	99%	3	100%
60S ribosomal protein L22 - Homo sapiens (Human)	15 kDa	2	100%	2	97%
60S ribosomal protein L11 - Homo sapiens (Human)	20 kDa	2	99%	1	77%
60S ribosomal protein L13 - Homo sapiens (Human)	24 kDa	2	97%	1	93%
40S ribosomal protein S19 - Homo sapiens (Human)	16 kDa	1	49%	2	85%
60S ribosomal protein L27a - Homo sapiens (Human)	17 kDa	1	96%	2	92%
40S ribosomal protein S8 - Homo sapiens (Human)	24 kDa	1	91%	1	87%
40S ribosomal protein S23 - Homo sapiens (Human)	16 kDa	1	94%	1	91%
ATP synthase subunit O, mitochondrial precursor - Homo sapiens (Human)	23 kDa	1	97%	1	91%
60S ribosomal protein L24 - Homo sapiens (Human)	18 kDa	1	88%	1	92%
Keratin, type I cytoskeletal 10 - Homo sapiens (Human)	60 kDa	10	100%	7	100%
Keratin, type I cytoskeletal 9 - Homo sapiens (Human)	62 kDa	7	100%	5	100%
Alpha-S2-casein precursor [Contains: Casocidin-1 - Bos taurus (Bovine)	26 kDa	5	100%	2	98%
Ig gamma-2 chain C region - Homo sapiens (Human)	36 kDa	3	99%	1	95%
Ig alpha-1 chain C region - Homo sapiens (Human)	38 kDa	1	64%	5	100%
Immunoblobulin light chain - Homo sapiens (Human)	24 kDa	1	100%	2	100%
Calcium/calmodulin-dependent protein kinase type II delta chain - Homo sapiens (Human)	56 kDa	1	91%	1	94%
Serum albumin precursor - Bos taurus (Bovine)	69 kDa	1	89%	1	70%

Table 9 Reproducible identifications of proteins co-purifying with TAP-HPH2a

This table includes LC/MS/MS identifications that were obtained from both TAP-HPH2a purifications (P1 and P2) conducted. The identifications were categorised as putative specific interaction (no highlighting), possible sepharose contaminant (light yellow highlighting) or contaminant (yellow highlighting), and sorted by purification

and number of identified peptides. In addition, the probabilities of correct identification are indicated per purification. Proteins in blue font indicate manual identification.

Identified Proteins	MW	P1	P2
E3 ubiquitin-protein ligase UBR5 - Homo sapiens (Human)	309 kDa	13	
Transferrin receptor - Homo sapiens (Human)	85 kDa	5	
Histone H2A - Nematostella vectensis (Starlet sea anemone)	20 kDa	4	
RNA binding motif protein 28 - Homo sapiens (Human)	86 kDa	4	
Dedicator of cytokinesis protein 7 - Homo sapiens (Human)	243 kDa	3	
Spectrin beta chain, brain 1 - Homo sapiens (Human)	275 kDa	3	
Uncharacterized protein C17orf85 - Homo sapiens (Human)	71 kDa	3	
ADP-ribosylation factor-like protein 6-interacting protein 4 - Homo sapiens (Human)	38 kDa	2	
Carboxypeptidase, vitellogenic-like - Homo sapiens (Human)	54 kDa	2	
E3 ubiquitin-protein ligase RING1 - Homo sapiens (Human)	42 kDa	2	
Heterochromatin protein 1-binding protein 3 - Homo sapiens (Human)	61 kDa	2	
NF-kappa-B-activating protein - Homo sapiens (Human)	47 kDa	2	
Pyruvate dehydrogenase - Homo sapiens (Human)	43 kDa	2	
RRP7A ribosomal RNA processing 7 homolog A - Homo sapiens (Human)	32 kDa	2	
Spectrin alpha chain, brain - Homo sapiens (Human)	285 kDa	2	
Spectrin beta chain, brain 2 - Homo sapiens (Human)	271 kDa	2	
60 kDa SS-A/Ro ribonucleoprotein - Homo sapiens (Human)	61 kDa	1	
Caspase 14 - Homo sapiens (Human)	28 kDa	1	
CDW4/GRWD1 - Homo sapiens (Human)	49 kDa	1	
Cell division protein kinase 10 - Homo sapiens (Human)	41 kDa	1	
CIR protein - Homo sapiens (Human)	50 kDa	1	
Coiled-coil-helix-coiled-coil-helix domain-containing protein 1 - Homo sapiens (Human)	13 kDa	1	
Cyclin-related protein FAM58A - Homo sapiens (Human)	28 kDa	1	
Far upstream element-binding protein 2 - Homo sapiens (Human)	73 kDa	1	
F-box only protein 22 - Homo sapiens (Human)	45 kDa	1	
Histone-binding protein RBBP4 - Homo sapiens (Human)	48 kDa	1	
Lipocalin-1 precursor - Homo sapiens (Human)	19 kDa	1	
Neuroguidin - Homo sapiens (Human)	36 kDa	1	
Nuclear protein NP60 - Homo sapiens (Human)	61 kDa	1	
Nuclease-sensitive element-binding protein 1 - Homo sapiens (Human)	36 kDa	1	
Nucleolar protein 10 - Homo sapiens (Human)	80 kDa	1	
P40 - Homo sapiens (Human)	40 kDa	1	
Pinin - Homo sapiens (Human)	82 kDa	1	
Probable rRNA-processing protein EBP2 - Homo sapiens (Human)	35 kDa	1	
Probable UPF0334 kinase-like protein C1orf57 - Homo sapiens (Human)	21 kDa	1	
Protein FAM133A - Homo sapiens (Human)	29 kDa	1	
Protein phosphatase from PCR H32 protein - Homo sapiens (Human)	3 kDa	1	
Protein S100-A8 - Homo sapiens (Human)	11 kDa	1	
Pumilio domain-containing protein KIAA0020 - Homo sapiens (Human)	74 kDa	1	
Putative helicase MOV-10 - Homo sapiens (Human)	114 kDa	1	
RPS9 protein - Homo sapiens (Human)	13 kDa	1	
SAFB2 protein - Homo sapiens (Human)	57 kDa	1	
Serine/arginine repetitive matrix 2 - Homo sapiens (Human)	191 kDa	1	
SMAD family member 9 - Homo sapiens (Human)	49 kDa	1	
TCOF1 protein - Homo sapiens (Human)	144 kDa	1	
Tetratricopeptide repeat protein 14 - Homo sapiens (Human)	88 kDa	1	
Transferrin - Homo sapiens (Human)	77 kDa	1	
Transformer-2 protein homolog - Homo sapiens (Human)	33 kDa	1	
U3 small nucleolar ribonucleoprotein protein IMP3 - Homo sapiens (Human)	22 kDa	1	
Uncharacterized protein C21orf70 - Homo sapiens (Human)	25 kDa	1	

Zinc finger protein 346 - Homo sapiens (Human)	33 kDa	1	
Splicing factor 3B subunit 2 - Homo sapiens (Human)	98 kDa		4
Polymerase delta-interacting protein 3 - Homo sapiens (Human)	46 kDa		3
Transcription elongation factor A protein-like 4 - Homo sapiens (Human)	25 kDa		2
DIS3 - Homo sapiens (Human)	109 kDa		2
Multiple myeloma tumor-associated protein 2 - Homo sapiens (Human)	29 kDa		1
Probable serine carboxypeptidase CPVL precursor - Homo sapiens (Human)	54 kDa		1
Numb-like protein - Homo sapiens (Human)	65 kDa		1
SNW domain-containing protein 1 - Homo sapiens (Human)	61 kDa		1
UPF0399 protein C6orf153 - Homo sapiens (Human)	30 kDa		1
Peroxisome biogenesis factor 1 - Homo sapiens (Human)	143 kDa		1
Centrosomal protein of 170 kDa - Homo sapiens (Human)	175 kDa		1
Putative splicing factor, arginine/serine-rich 14 - Homo sapiens (Human)	120 kDa		1
Mitochondrial fission regulator 1 - Homo sapiens (Human)	37 kDa		1
Liprin-beta-1 - Homo sapiens (Human)	114 kDa		1
MAP7 domain-containing protein 3 - Homo sapiens (Human)	98 kDa		1
Tubulin, beta polypeptide - Homo sapiens (Human)	48 kDa	8	
Matrin 3 - Homo Sapiens (Human)	100 kDa	6	
Heterogeneous nuclear ribonucleoproteins C1/C2 - Homo sapiens (Human)	34 kDa	5	
40S ribosomal protein S4, X isoform - Homo sapiens (Human)	30 kDa	5	
Heterogeneous nuclear ribonucleoproteins A2/B1 - Homo sapiens (Human)	37 kDa	4	
Nucleolar RNA helicase 2 - Homo sapiens (Human)	87 kDa	4	
Probable ATP-dependent RNA helicase DDX5 - Homo sapiens (Human)	69 kDa	4	
Protein-L-isoaspartate(D-aspartate) O-methyltransferase - Homo sapiens (Human)	25 kDa	4	
Tubulin alpha-1B chain - Homo sapiens (Human)	50 kDa	4	
Vimentin - Homo sapiens (Human)	54 kDa	4	
Ribosomal protein S3A - Homo sapiens (Human)	30 kDa	4	
RPL14 protein - Homo sapiens (Human)	24 kDa	4	
ATP synthase subunit O, mitochondrial precursor - Homo sapiens (Human)	23 kDa	3	
DEAH - Homo sapiens (Human)	141 kDa	3	
Peroxiredoxin-1 - Homo sapiens (Human)	22 kDa	3	
40S ribosomal protein S25 - Homo sapiens (Human)	14 kDa	3	
60S acidic ribosomal protein P0 - Homo sapiens (Human)	34 kDa	3	
60S ribosomal protein L26 - Homo sapiens (Human)	17 kDa	3	
60S ribosomal protein L4 - Homo sapiens (Human)	48 kDa	3	
Ribosomal protein L18 - Homo sapiens (Human)	19 kDa	3	
Ribosomal protein L7a - Homo sapiens (Human)	22 kDa	3	
Heterogeneous nuclear ribonucleoprotein AB isoform a variant - Homo sapiens (Human)	36 kDa	2	
Heterogeneous nuclear ribonucleoprotein K - Homo sapiens (Human)	51 kDa	2	
Myosin light chain kinase 2, skeletal/cardiac muscle - Homo sapiens (Human)	65 kDa	2	
Probable ATP-dependent RNA helicase DDX27 - Homo sapiens (Human)	90 kDa	2	
DnaJ - Homo sapiens (Human)	254 kDa	2	
Heat shock 70kDa protein 9B - Homo sapiens (Human)	74 kDa	2	
Heat shock protein HSP 90-beta - Homo sapiens (Human)	83 kDa	2	
28S ribosomal protein S31, mitochondrial precursor - Homo sapiens (Human)	45 kDa	2	
40S ribosomal protein S11 - Homo sapiens (Human)	18 kDa	2	
40S ribosomal protein S14 - Homo sapiens (Human)	16 kDa	2	
40S ribosomal protein S18 - Homo sapiens (Human)	18 kDa	2	
40S ribosomal protein S20 - Homo sapiens (Human)	13 kDa	2	
60S ribosomal protein L38 - Homo sapiens (Human)	8 kDa	2	
Ribosomal protein L12 variant - Homo sapiens (Human)	21 kDa	2	
RPL3 protein - Homo sapiens (Human)	34 kDa	2	

Small nuclear ribonucleoprotein Sm D3 - Homo sapiens (Human)	14 kDa	1	
Actin, alpha skeletal muscle - Homo sapiens (Human)	42 kDa	1	
ELAV-like protein 1 - Homo sapiens (Human)	36 kDa	1	
Elongation factor 1-alpha 2 - Homo sapiens (Human)	50 kDa	1	
Heterogeneous nuclear ribonucleoprotein A1 - Homo sapiens (Human)	39 kDa	1	
Heterogeneous nuclear ribonucleoprotein U-like protein 2 – Homo sapiens (Human)	85 kDa	1	
MCM7 minichromosome maintenance deficient 7 - Homo sapiens (Human)	81 kDa	1	
RNA-binding motif protein, X-linked-like-2 - Homo sapiens (Human)	43 kDa	1	
Small nuclear ribonucleoprotein Sm D3 - Homo sapiens (Human)	14 kDa	1	
Thioredoxin - Homo sapiens (Human)	9 kDa	1	
Short heat shock protein 60 Hsp60s2 - Homo sapiens (Human)	27 kDa	1	
40S ribosomal protein S10 - Homo sapiens (Human)	19 kDa	1	
40S ribosomal protein S17 - Homo sapiens (Human)	16 kDa	1	
40S ribosomal protein S2 - Homo sapiens (Human)	31 kDa	1	
40S ribosomal protein S6 - Homo sapiens (Human)	29 kDa	1	
60S acidic ribosomal protein P2 - Homo sapiens (Human)	12 kDa	1	
60S ribosomal protein L10 - Homo sapiens (Human)	25 kDa	1	
60S ribosomal protein L19 - Homo sapiens (Human)	23 kDa	1	
60S ribosomal protein L29 - Homo sapiens (Human)	18 kDa	1	
60S ribosomal protein L30 - Homo sapiens (Human)	13 kDa	1	
60S ribosomal protein L31 - Homo sapiens (Human)	14 kDa	1	
60S ribosomal protein L7 - Homo sapiens (Human)	29 kDa	1	
Peroxiredoxin 6 - Homo sapiens (Human)	9 kDa	1	
Ribosomal protein S19 binding protein 1 - Homo sapiens (Human)	15 kDa	1	
Stress-70 protein, mitochondrial precursor - Homo sapiens (Human)	74 kDa		4
Tubulin beta chain - Homo sapiens (Human)	50 kDa		3
60S ribosomal protein L3 - Homo sapiens (Human)	46 kDa		2
HEJ1 - Homo sapiens (Human)	7 kDa		1
LIM domain and actin-binding protein 1 - Homo sapiens (Human)	85 kDa		1
Immunoblobulin G1 Fab heavy chain variable region - Homo sapiens (Human)	25 kDa		1
60S ribosomal protein L12 - Homo sapiens (Human)	18 kDa		1
40S ribosomal protein S19-binding protein 1 - Homo sapiens (Human)	15 kDa		1
28S ribosomal protein S7, mitochondrial precursor - Homo sapiens (Human)	28 kDa		1
Keratin 1 - Pan troglodytes verus	65 kDa	14	
Alpha-S1-casein precursor [Contains: Antioxidant peptide] - Bos taurus (Bovine)	25 kDa	5	
IGHG1 protein - Homo sapiens (Human)	51 kDa	5	
Keratin, type II cytoskeletal 2 epidermal - Homo sapiens (Human)	66 kDa	5	
Genome polyprotein [Contains: P1 proteinase - Tobacco etch virus (TEV)	346 kDa	4	
Immunoglobulin lambda protein - Homo sapiens (Human)	25 kDa	4	
Hair acidic keratin 1 - Capra hircus (Goat)	47 kDa	3	
Poly(rC)-binding protein 1 - Bos taurus (Bovine)	37 kDa	2	
Ig kappa chain V-I region EU - Homo sapiens (Human)	13 kDa	2	
KRT5 protein - Bos taurus (Bovine)	63 kDa	2	
Lysozyme C precursor - Gallus gallus (Chicken)	16 kDa	2	
MGC139925 protein - Bos taurus (Bovine)	54 kDa	2	
LOC733345 protein - Xenopus laevis (African clawed frog)	27 kDa	1	
Lysozyme C precursor - Homo sapiens (Human)	17 kDa	1	
Putative ATP-dependent RNA helicase DHX30 - Bos taurus (Bovine)	136 kDa	1	
RPS4Y1 - Pongo pygmaeus (Bornean orangutan)	2 kDa	1	
Beta-casein - Capra hircus (Goat)	21 kDa	1	
Keratin associated protein 11-1 - Homo sapiens (Human)	17 kDa	1	
Serum albumin precursor - Bos taurus (Bovine)	69 kDa	1	

Keratin, type II cytoskeletal 1 - Homo sapiens (Human)	66 kDa		7
Beta-lactoglobulin precursor - Bos taurus (Bovine)	20 kDa		1
Trypsinogen 10 - Mus musculus (Mouse)	26 kDa		1
Ig lambda chain C regions - Homo sapiens (Human)	11 kDa		1
Calmodulin - Homo sapiens (Human)	17 kDa		1

Table 10 Single identifications of proteins co-purifying with TAP-HPH2a

This list includes LC/MS/MS identifications that were obtained from only one of the two TAP-HPH2a purifications (P1 and P2) conducted. The identifications were categorised as putative specific interaction (no highlighting), possible sepharose contaminent (light yellow highlighting) or contaminent (yellow highlighting), and sorted by purification and number of identified peptides. Proteins in blue font indicate manual identification.

Figure 6.4 Sense/antisense-specific RT-qPCR Methodology

For a given primer pair that is used to specifically amplify genomic DNA, such as the *INK4A* promoter, the individual primers also provide complementary binding to the transcripts sense or antisense to this region (step 1). Therefore such such an individual primer can be used to specifically reverse transcribe the transcripts either sense or antisense to this region (step 2). Consequently the cDNA is purified and now both primers of the pair are used to amplify the cDNA in a qPCR reaction (step 3).

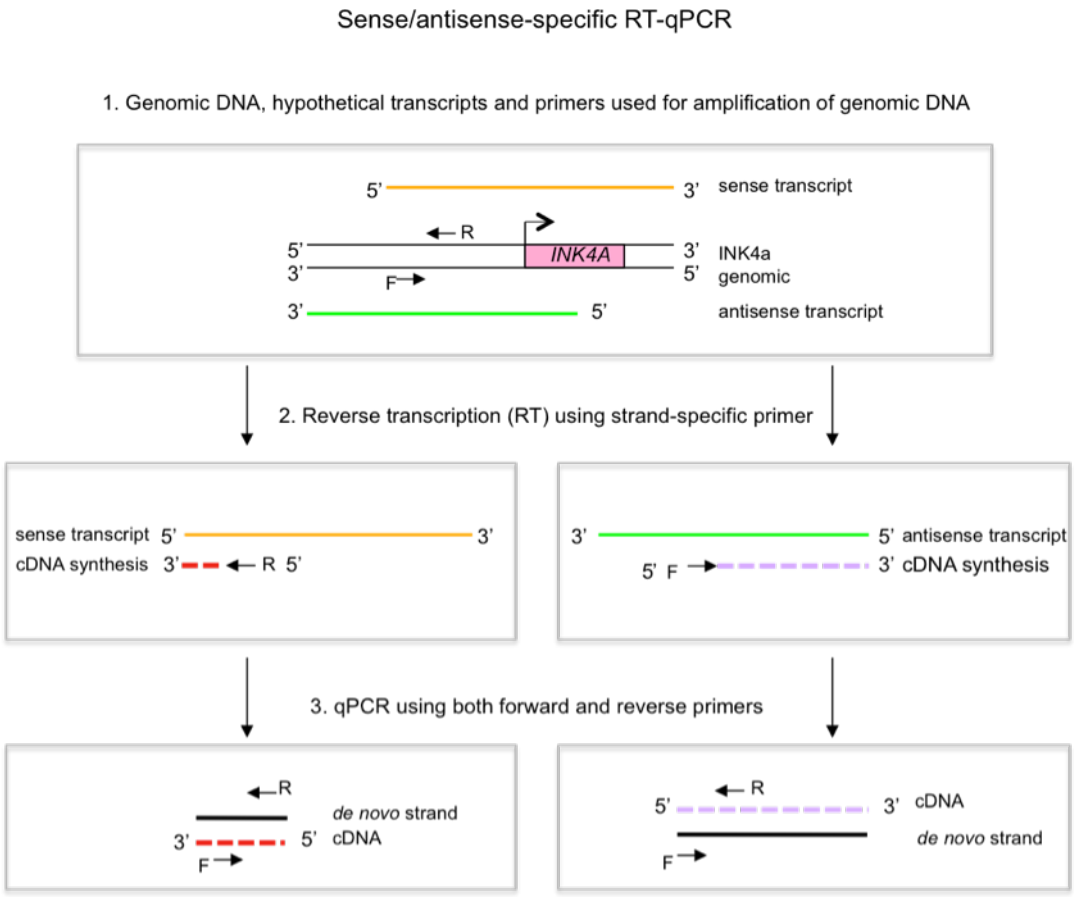


Figure 6.5 Profile of *INK4B-ARF-INK4A* sense and antisense transcription

A) Sense and antisense-specific RT reactions were performed with specific primers along the locus and using 0.125µg of total RNA from IMR-90 (PD18) (as described in section **2.5.4**). cDNA was purified and interrogated by qPCR with the matching forward and reverse primer set. Resulting amplification Ct values were normalised for relative primer efficiency and divided by a constant c=0.05 to provided simplified comparison and plotted.

B) Sense and antisense-specific RT-qPCR as described under A) using 0.100µg of total RNA from mid-lifespan Hs68 cells.

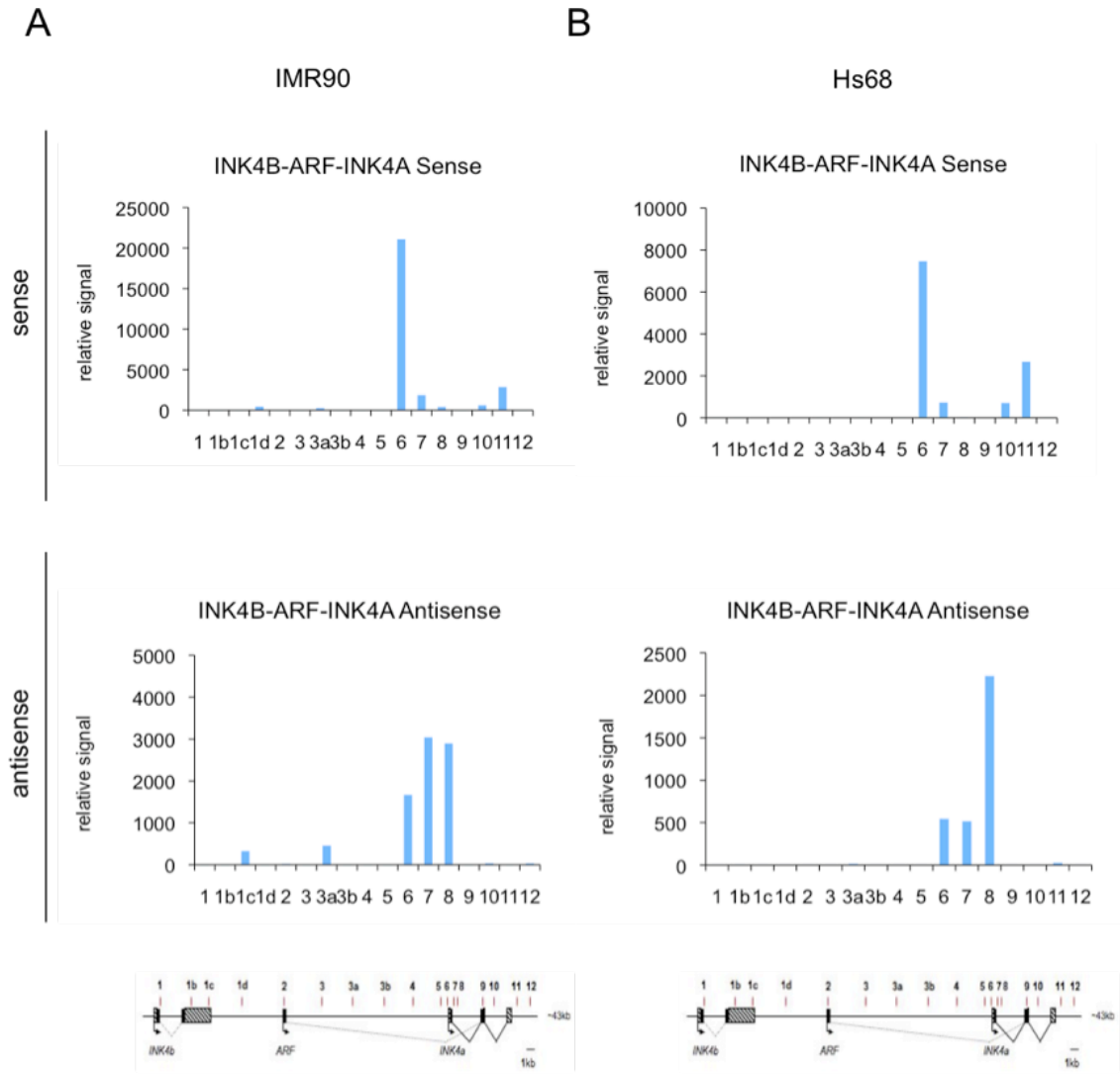


Figure 6.6. HPH2 co-immunoprecipitates RNA transcripts antisense to the *INK4A* promoter region

RNA ChIP was performed with unspecific IgG (Abcam), CBX7 (Abcam) and HPH2 (1615) (as described in section **2.5.5**). Purified RNA was reverse transcribed using primer 6-F and PCR was performed using the PS6 primer pair (described in the Materials and Methods section). A control RT reaction was performed for each IP without the reverse transcriptase, but no amplification was observed for these samples in qPCR and thus these have not been plotted. Error bars indicate standard deviation between unrelated biological replicates, with n=2.

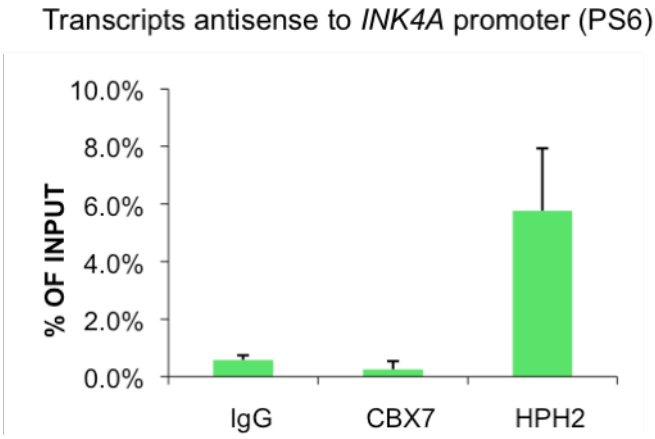
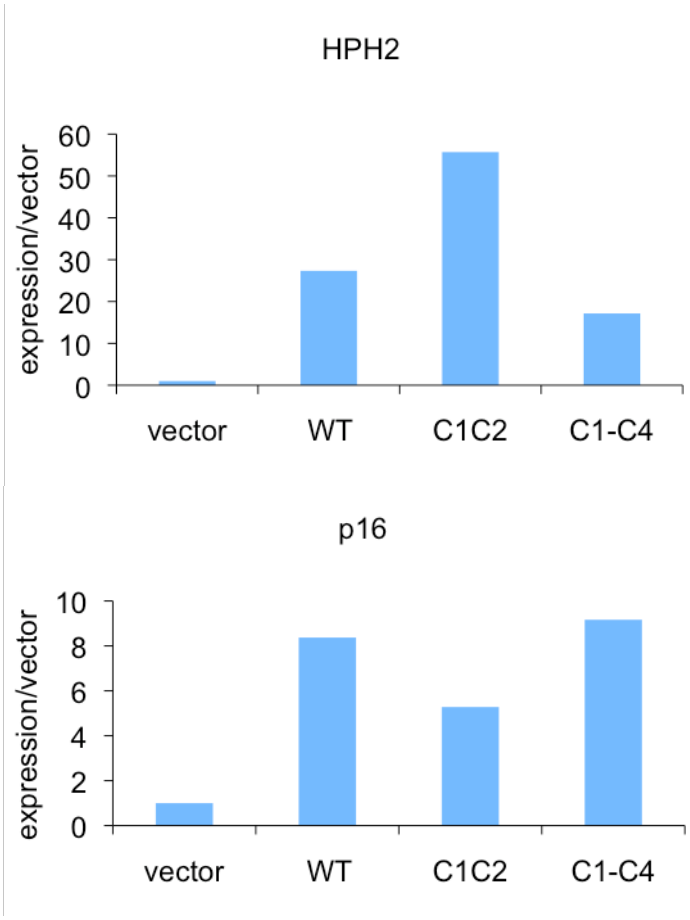


Figure 6.7 Mutation of the FCS RNA binding domain of HPH2 does not affect its induction of *INK4A*

Hs68 cells (PD33) were infected with pBABE pu encoding wildtype HPH2 or the C1C2 or C1-C4 mutants, or vector control. 5 days following selection cells were harvested and RNA was extracted (as described in section 2.5.1). RT-qPCR was performed to measure the transcript levels of p16, HPH2 and GAPDH (internal standard). Resulting transcript levels were normalised to the pBABE vector control.



7 Discussion

This thesis set out to define the involvement of the HPH family proteins in the repression of *INK4A* and a number of achievements to this end have been described. Firstly, a series of antibodies against human CBX7 and the HPH proteins were generated and evaluated. Secondly, it was shown that contrary to the preliminary indications, there is no *in vitro* evidence for specific interactions between individual members of the HPH and CBX families although the possibility remains open for a preferential interaction between CBX7 and HPH2 *in vivo*. Thirdly, all three HPH proteins were shown to be involved in the repression of *INK4A* in primary human fibroblasts and for HPH2 (and HPH3, data not shown) it was confirmed that they are directly associated with the locus. Curiously, overexpression of HPH2 was found to activate expression of *INK4A*, typically associated with loss of PcG-mediated repression by an as yet uncharacterised mechanism. Fourthly, to gain a better understanding of the contribution of the HPH family to PcG-mediated repression, the proteins interacting with HPH2 were identified by affinity purification and mass spectrometry, revealing a large number of proteins implicated in RNA processing. Lastly, HPH2 was shown to interact with an RNA antisense to the *INK4A* promoter.

7.1 Lessons about PRC1 composition

The fact that mammalian cells express multiple paralogues of each of the *Drosophila* PRC1 components presents a considerable challenge as well as a fascinating topic for research. A comprehensive study of the individual core proteins would in principle require a minimum of 16 antibodies, notwithstanding possible splice variants and post-translational modifications, and obtaining suitable reagents for immunodetection and chromatin immunoprecipitations continues to be a major obstacle for progress in the field. Difficulties in expressing the recombinant proteins and cross-reactivity against conserved domains, as discussed in Chapter 3, add to this challenge.

In considering why there are so many PcG paralogues in mammalian cells, a central question is whether there are any preferential interactions between individual members of each family. In the literature, the PRC1 complex is often described as a single entity. Purifications of PRC1 complexes rather have provided mixed messages

about stoichiometry and composition. Parallel work in our laboratory has demonstrated convincingly that each PRC1 complex in human cells interacts with only a single representative of the Pc, Psc, Ph or Sce families (Maertens et al., 2009). The characterization of HPH2 interactions described in Chapter 6 supports this contention in that there was no evidence for co-purification of HPH1 or HPH3. It is nevertheless clear that all three HPH proteins can be co-purified with BMI1, MEL18 and CBX8 (Dietrich et al., 2007; Maertens et al., 2009) and that CBX2 and CBX8, and both RING proteins co-purify with HPH2. This argues against specificity between particular members of the PRC1 subfamilies and in direct comparisons, all five human Pc proteins were shown to interact with HPH2 when transiently co-expressed (Chapter 4). Although the data by no means exclude the possibility of preferential associations *in vivo*, perhaps based on relative expression levels in different cell types, collectively they suggest that there are multiple permutations of PRC1 complexes.

Based on their common ancestry, it might be expected that the several paralogues of each PRC1 core component have the same or similar functions. This could mean that depending on their relative abundance, they are used interchangeably in the assembly of a functional PRC1 complex without any particular logic. However, a number of reports ascribe specific functions to particular PRC1 proteins or implicate these proteins in unique mechanisms that most likely cannot be generalised. Examples of such unqiue mechanisms include the SUMO E3 ligase activity of CBX4 or unique putative ATM/ATR phosphorylation sites in HPH2 and HPH3. While it seems unlikely that every homologue provides unique functionality or levels of regulation, differences in the expression of PcG proteins in different cell types may be explained by some degree of functional diversity tailored to specific needs.

Another way to rationalise the multiplicity of PRC1 complexes would be if they had preferred target genes, perhaps dictated by specific elements that are not part of the core PRC1 complex. However, one of the more remarkable findings in this work was that all three human HPH proteins appear to contribute to the regulation of the *INK4a* gene in human fibroblasts, as judged by shRNA-mediated knockdown. Recently published work from our laboratory and others (Cales et al., 2008; Dietrich et al., 2007; Gil et al., 2005; Jacobs et al., 1999a; Maertens et al., 2009) has drawn similar conclusions for two Pc orthologues (CBX7 and CBX8), two Psc proteins (BMI1 and MEL18) and at least one of the Sce proteins (RING1B). These findings are also supported by genetic ablation of PcG genes in mice, which in several cases has been reported to cause derepression of *Ink4a/Arf* and premature senescence in the derived

MEFs (Cales et al., 2008; Isono et al., 2005a; Jacobs et al., 1999a). It thus appears that so far no PRC1 component has been characterised that does definitely not link to the regulation of *INK4a*.

After having argued that each PRC1 complex comprises of only a single representative from the Pc, Psc, Ph and Sce families, the involvement of HPH1, HPH2 and HPH3 effectively triples the number of PRC1 complexes that are implicated in the regulation of *INK4a*. In fact, these results are the first demonstration that all homologues of a particular PRC1 core component are implicated in the regulation of *INK4A* (if not any PcG target gene). Simplistically, this number of PRC1 complexes regulating *INK4A* currently stands at 12 (2Pc x 2Psc x 3Ph) and if both RING proteins are involved it will rise to 24. The rationale is very puzzling and suggests that the simple paradigm of a single generic PRC1 complex binding to a target gene is untenable. Importantly, each of the tested proteins shows the same binding profile across the promoter and first exon of *INK4A*, and ChIP-reChIP analyses with CBX7 and CBX8 indicated that they are bound to the same piece of DNA (Maertens et al., 2009), although the data does not exclude the possibility of cross-linking between separate alleles of *INK4A*. The generation and validation of ChIP grade antibodies against HPH2 (and HPH3) as described in this thesis now open up the possibility to do a similar assessment of the binding of two different HPH proteins to a single copy of *INK4A*.

7.2 Interactions between PcG complexes

The fact that depletion of either HPH1, HPH2 or HPH3, or the different Pc and Psc homologues, leads to loss of repression of *INK4A* implies that they are not functionally redundant. However, it is difficult to come up with an explanation based on current understanding of the mechanism. One important observation is that the shRNA-mediated knockdown of CBX7 results in the depletion of both CBX7 and CBX8 at the *INK4A* locus as judged by ChIP, and vice versa (Maertens et al., 2009). The same conclusions apply to the two PSC proteins, BMI1 and MEL18, leading to the idea that the PRC1 proteins bind to *INK4A* as a complex of complexes. In this scenario, very much like a 'house of cards', loss of one component can lead to the collapse of the entire system and it will be interesting to extend the observations to the HPH proteins. The demonstrated oligomerisation potential of the HPH proteins could play a fundamental role in the formation of these higher order structures (Kim et al., 2002).

Hence the HPH proteins could present particular weak spots in maintaining the overall structure of the complex and this could be further characterised by the effect of recently generated SAM oligomerisation mutants in. It thus remains to be determined what the structure would be of a PRC1 multicomplex and how polyhomeotic may structurally contribute to it.

Another unknown in the context of PcG-mediated repression is the extent to which the PRC1 and PRC2 complexes co-operate. The two complexes are often co-localised at repressed genes and it has been proposed that a transient interaction is required for setting up PcG-mediated repression (van Lohuizen et al., 1998). The identification of RBBP4/NURF55 as a putative binding partner of HPH2 would be consistent with this idea but needs to be confirmed. Another element that might link the two complexes is the interaction with RNA. As the results in Chapter 6 provide a preliminary basis for the interaction of HPH2 with RNA antisense to *INK4A*, the use of ncRNAs might be shared by both PRC2 and PRC1 and may add directionality to PcG repression.

7.3 The HPH proteins and cell proliferation

Although the complex-of-complexes model serves to fit the discussed observations, it is difficult to reconcile with the effects observed upon over-expression of particular PRC1 components. For example, it is well documented that ectopic expression of Bmi1 and Cbx7 can delay replicative senescence in human fibroblasts through the repression of *INK4A* (Bracken et al., 2007; Gil et al., 2004). This fits well with the evidence that both *Bmi1* and *Cbx7* function as oncogenes in transgenic mice (Jacobs et al., 1999b; Scott et al., 2007a). Indeed, Bmi1 was initially identified by its ability to collaborate with Myc in retrovirally induced lymphomas in mice (Haupt et al., 1991; van Lohuizen et al., 1991). The oncogenic effects can largely be explained in terms of suppression of mouse *Ink4a/Arf*. Conversely, ablation of Bmi1 leads to loss of proliferative potential due to derepression of *Ink4a/Arf.* A similar line of reasoning would implicate the mouse *HPH2* homologue (*Mph2*) as a potential oncogene (Isono et al., 2005a).

The question that arises is whether the ability of BMI1 or CBX7 to repress *INK4A* is dependent on HPH2 or other HPH proteins. The ‘house of cards’ model implies that the PRC1 complexes are interdependent but knockdown of HPH2 did not prevent mCbx7-mediated repression of *INK4a* (Chapter 5). This could both imply that

HPH2 is not rate limiting or that its functions can be substituted by HPH1 or HPH3. Similarly, it was reported that the ability of mouse Cbx7 to repress mouse *Ink4a/Arf* was not dependent on *Bmi1* (Gil et al 2004), again contradicting the model. However, in the MEF experiments conducted in this work, *Mph2* did appear to be important for mCbx7-mediated effects. Clearly, further work will be required to resolve these inconsistencies and the tools for including HPH3 in the analyses are now in place.

Further work will also be required to address the apparent anachronism that over-expression of HPH2 activates rather than represses *INK4a*. This cannot be dismissed as an oddity, as similar findings have been reported for MEL18 ((Guo et al., 2007a; Guo et al., 2007b; Tetsu et al., 1998)) and with CBX8 (Julie Stock, personal communication). It is conceivable that the over-expressed proteins are acting in a dominant-negative fashion by sequestering or titrating out a rate limiting factor or activity, but this would have to be a specific mechanism for some, but not all PRC1 proteins, as BMI1 or CBX7 overexpression has a diametrically opposite effect. Alternatively, over-expressed HPH2 might be acting indirectly by suppressing the expression of a protein or several proteins that are essential for repression of *INK4a.* Obvious candidates would be the PcG genes themselves. Also, given the precedents for autoregulation of the *Drosophila* polyhomeotic genes, it is important to consider the potential intra-regulation of polyhomeotic genes (Bloyer et al., 2003; Fauvarque et al., 1995). It has also been reported that *CBX4, CBX7, CBX8* and *BMI1* may be subject to PcG repression (Bracken et al., 2006), and recently our laboratory found through ChIP-Seq that CBX7 and CBX8 co-localise on the CBX7 promoter. For this to be a tenable explanation the extent to which PcG genes are regulated by PRC1 complexes would have to be explored in more depth and a distinction would have to be drawn between the PcG genes that are subject to feedback regulation and those that are not.

As well as cross-regulation of PcG genes, the effects of HPH overexpression both on *INK4A* expression and cell proliferation could involve a wide array of other PcG target genes. For example, a recent report demonstrated that polyhomeotic has a role in mediating control of proliferation by regulating Notch signalling in *Drosophila*, therefore ascribing it a 'tumour suppressor' role (Martinez et al., 2009). A tumour suppressive function is also supported by the observed loss of expression of HPH homologues in human cancer (Deshpande et al., 2007; Tokimasa et al., 2001). It will therefore be important to study the overexpression of each HPH protein in order to better understand their effects. In this context, it might be informative to compare gene

expression profiles in human fibroblasts in which each HPH protein has been either knocked down or over-expressed.

7.4 Novel interacting partners for HPH2

Whereas PC proteins have chromodomains that are implicated in binding to the H3K27me3 mark and the PSC-SCE dimers have RING finger domains that catalyse ubiquitination of H2AK119, little is known about the functional contribution of PH to the PRC1 complex. It was hoped that the tandem affinity purification of HPH2 complexes would shed light on non-PcG proteins that are recruited by binding to HPH2, as well as extract possible differences between the HPH2a and HPH2b isoforms. Although the preliminary findings looked very promising, it proved difficult to scale up the purification to a level that was amenable to mass spectrometry. This prevented a proper comparison between HPH2a and HPH2b within the span of this work, but in principle, the longer isoform demonstrated a greater potential for identifying novel partners. It is debatable whether efforts should be made to improve the purification protocol for improved yield, for example by altering the expression vector or seeking transient rather than stable expression, or rather to focus on the validation of the current list of candidate proteins. Repeating the purification with the short isoform or one of the other HPH proteins has the potential to uncover common interactions shared by all HPH proteins and simplify the follow up.

It will be important to assess whether any of the novel interacting proteins are associating with the PRC1 complex. Previous reports suggest that PRC1 proteins can act outside of the context of the classical PRC1 complex, often associated with specific functions (reviewed in (Simon and Kingston, 2009)). An example of this is the RING-associated factor (dRAF) complex in *Drosophila* or the BCL6 co-repressor (BCOR) complex in mammals, which is capable of performing efficient ubiquitylation of H2A without the full complement of PRC1 components (Gearhart et al., 2006; Lagarou et al., 2008). No HPH proteins were present in these complexes and in reconstitution experiments it has proved possible to generate active complexes without a HPH protein (Levine et al., 2002). Such unconventional functionality is not restricted to PRC1, as the PRC2 complex was shown to have a distinct cytosolic function in the regulation of actin polymerisation (Su et al., 2005). These findings suggest that PcG complexes or derivatives thereof can be involved in distinct functions.

HPH proteins therefore could also carry out independent of conventional PRC1 functions. As such, it was reported that HPH1 is distinct from other PRC1 core components in timing of recruitment to the inactive X-chromosome (Plath et al., 2004). Furthermore, apart from known PRC1 proteins, none of the potential HPH2-associated proteins identified by mass spectrometry featured in the affinity purified CBX7, BMI1 and MEL18 complexes characterized previously in our laboratory, although these complexes did contain HPH2. This opens up the possibility that some of the functions attributed to HPH2 in this thesis may reflect its participation in separate non-conventional PRC1 or even independent complexes.

7.4.1 Splicing and RNA binding

The finding that HPH2 is associated with multiple RNA processing proteins, many of them involved in splicing, suggests that HPH2 is closely associated with RNA. It is becoming increasingly evident that the processes of chromatin regulation, transcription and splicing are tightly integrated and there is a clear precedent for the recruitment and regulation of splicing factors by chromatin modifiers (reviewed in (Allemand et al., 2008)). Recent literature suggests that splicing occurs directly at the level of the nascent transcripts and thereby mediates the production of transcript variants for nearly all mammalian genes (Listerman et al., 2006; Pandya-Jones and Black, 2009; Wang et al., 2008). The involvement of chromatin and its modifiers in this process has recently gained further support as it was found that nucleosomes are preferentially localised at exonic regions and may play an important role in exon recognition (Schwartz et al., 2009; Tilgner et al., 2009). This would suggest that processes involved in PcG-mediated repression that either modify the nucleosomes or alter their compaction may potentially influence the effectiveness of splicing.

Previous findings have mainly been centred around a connection between splicing and the chromatin remodelling activity of SWI/SNF in the context of active transcription, but a link between PcG-mediated repression and splicing has also previously been established (Isono et al., 2005b). It will be interesting to define how these opposing systems modulate splicing and if the PcG proteins have indeed a negative effect on splicing. The potentially negative involvement of HPH2 in this process could produce an instant effect on the transcription of target genes before promoter silencing is achieved. Furthermore, alternative splicing is influenced by the

modification of nucleosomes, which can create a barrier for transcriptional elongation (de la Mata et al., 2003; Hodges et al., 2009). This would provide a putative mechanism for the PRC1 complex to block transcriptional elongation or by preventing the processing of primary transcripts, and may possibly collaborate with the effects of histone H2A ubiquitination.

Another possibility would be that HPH2 is closely involved in the processing of ncRNAs employed by the PcG system for targeting specific genes (reviewed in (Simon and Kingston, 2009)). Rather than being an exclusive property of PRC2 complexes, as previously implied, the RNA binding capacity of HPH2 and the RNA processing activities of its interaction partners might function in handling transcripts that will guide the PRC complexes to their targets. Alternatively, the ncRNA might have a direct role in gene silencing and will have to be processed at the chromatin of the target gene. There is previous evidence suggesting that ncRNA contributes to gene silencing by infiltrating the DNA double strand, leading to the formation of a triplex structure (Martianov et al., 2007). Although these observations have not been connected to PcG-mediated repression, this might be a general mechanism by which ncRNA contributes to sequence specific gene silencing.

It is therefore encouraging to find HPH2 binding to RNA that is antisense with respect to the *INK4A* promoter. Although these preliminary findings need to be further investigated, there is precedence for antisense transcription through the *INK4B-ARF-INK4A* locus (Gonzalez et al., 2008; Pasmant et al., 2007). In addition, the low resolution transcriptional profiling of the *INK4B-ARF-INK4A* locus performed highlights some interesting transcriptional activity that appears to localise with PcG binding. The laboratory is currently undertaking efforts to map the strand-specific transcription of the locus at a much higher resolution. Nevertheless, a common principle about how ncRNA is involved in PcG repression is still lacking, which limits any such strategies. In this regard the finding that a substantial number of conserved long ncRNAs interact with PRC2 is encouraging (Khalil et al., 2009), but it is not yet clear whether these RNA molecules consistently work *in trans* to recruit PRC2, or work *in cis* analogous to the PcG recruitment by *XIST* (Rinn et al., 2007; Zhao et al., 2008). The list of ncRNAs reported by Khalil and colleagues did not include any transcripts spanning the *INK4A-ARF-INK4B* locus.

These highly speculative ideas will need to be addressed carefully. The current priority lies in the identification of the RNA molecules bound by HPH2. This would be best addressed by employing high-throughput sequencing of reverse transcribed RNAs

using technology that produces long reads (such as the 454 sequencer from Roche). This is imperative as there is no reference sequence for the full transcriptome and substantial variability in transcripts would hamper RNA sequence reconstruction based on short sequence reads. The identified RNA molecules could then be matched back to the sequence of *INK4/ARF* to find sense/antisense sequence alignments. The matching transcripts could then be further investigated through modulation of their expression levels and studying the regulation of their endogenous expression.

7.4.2 Higher-order chromatin regulation

The involvement of PRC1 complexes in ~~a~~ higher order chromatin structure has previously been demonstrated for the regulation of the homeotic bithorax gene cluster in *Drosophila* (Lanzuolo et al., 2007). This concept might not be restricted to repression of gene clusters, but could extend to single target genes. It has been proposed that intra-locus chromatin looping may be a general feature of PcG repression that is mediated by the H3K27me3 mark (Schwartz and Pirrotta, 2007; Simon and Kingston, 2009). Intra- and interchromosomal contacts between such loops could establish nuclear domains that impose repression on large collections of target genes (Tiwari et al., 2008).

The H1 linker histone is involved in the compaction of the chromatin fiber and therefore is involved in the establishment of higher order structures. The co-purification of histone H1.2 (H1d) with HPH2a thus provides an interesting prospect, although it will first be necessary to determine whether this represents a direct interaction between the two proteins. As previously mentioned, one of the isoforms of EED enables the PRC2 complex to trimethylate histone H1 on K26, which is conserved in most variants of H1 (Kuzmichev et al., 2004). In fact, the sequence surrounding H1K26 is highly similar to that of H3K27, but only in isoform H1.4 (H1b) (reviewed in (Schwartz and Pirrotta, 2007)). It thus remains to be determined if H1.2 is also a target of PRC2 methylation in the context of *INK4A* repression and whether this modification has any impact on the putative interaction with HPH2.

7.4.3 DNA damage response

Another fascinating angle that deserves further investigation is the possible role of HPH2 in the DNA damage response (DDR). This is of particular interest, following the recent identification of a role for Bmi1 in regulating the DDR, as most of the *Bmi1*$^{-/-}$ phenotype could be rescued by knockout of *Chk2* (Liu et al., 2009). Furthermore, it has been reported that the strength of the DDR is determined by overall chromatin compaction, which is limiting access to the double strand breaks (Fernandez-Capetillo and Murga, 2008). Removal of PRC1 from repressed loci could allow de-compaction of the chromatin fiber and access to the sites of DNA damage. The putative interaction between HPH2 and histone H1.2 could further contribute to this process, as a 50% reduction of H1 levels has been shown to increase the magnitude of the DDR (Murga et al., 2007). This study reports that a similar increase could also be obtained by inhibition of HDAC activity by trichostatin A, which would be in line with a reduction of PcG activity.

Although the functional relevance has not been explored, HPH2 and HPH3 both have potential target sites for phosphorylation by ATM or ATR. This also applies to a number of the proteins that co-purify with HPH2, including RBM14/CoAA, heterochromatin protein 1-binding protein 3 (HP1BP3/H1BP74) and others (Matsuoka et al., 2007). It will therefore be interesting to develop phospho-specific antibodies against HPH2 and HPH3 to study whether it indeed becomes modified upon DNA damage and how this affects its binding to *INK4A*. The identification of the ubiquitin ligase EDD in the HPH2 complex further strengthens the case for a link with the DDR. EDD is known to interact with the damage-specific DNA binding protein 1 (DDB1). Interestingly, DDB1 is also implicated, along with CUL4, as a component of a Trithorax like complex that is required for RAS-induced activation of *INK4A* (Kotake et al., 2009; Maddika and Chen, 2009). A high priority for the future will be to determine whether EDD directly interacts with HPH2 and if so through which domains.

7.5 Conclusions

This thesis has addressed the suspected role of the HPH proteins in the regulation of *INK4A* and has sparked a number of different avenues to pursue in order to understand

the impact of the HPH proteins on *INK4A*-related complex composition, the interaction between these complexes, a putative role for an RNA component, and additional functions that might illuminate novel aspects of the PcG-mediated repression mechanism. It is hoped that this work will stimulate further interest in the HPH proteins and can serve as a foundation for exploring these exciting new avenues.

References

- Acosta, J.C., A. O'Loghlen, A. Banito, M.V. Guijarro, A. Augert, S. Raguz, M. Fumagalli, M. Da Costa, C. Brown, N. Popov, Y. Takatsu, J. Melamed, F. d'Adda di Fagagna, D. Bernard, E. Hernando, and J. Gil. 2008. Chemokine signaling via the CXCR2 receptor reinforces senescence. Cell. 133:1006-18.
- Agger, K., P.A. Cloos, L. Rudkjaer, K. Williams, G. Andersen, J. Christensen, and K. Helin. 2009. The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence. Genes Dev. 23:1171-6.
- Akasaka, T., M. Kanno, R. Balling, M.A. Mieza, M. Taniguchi, and H. Koseki. 1996. A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. Development. 122:1513-22.
- Akasaka, T., K. Tsuji, H. Kawahira, M. Kanno, K. Harigaya, L. Hu, Y. Ebihara, T. Nakahata, O. Tetsu, M. Taniguchi, and H. Koseki. 1997. The role of mel-18, a mammalian Polycomb group gene, during IL-7-dependent proliferation of lymphocyte precursors. Immunity. 7:135-46.
- Akhtar, A., D. Zink, and P.B. Becker. 2000. Chromodomains are protein-RNA interaction modules. Nature. 407:405-9.
- Al-Hajj, M., M.S. Wicha, A. Benito-Hernandez, S.J. Morrison, and M.F. Clarke. 2003. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 100:3983-8.
- Alkema, M.J., M. Bronk, E. Verhoeven, A. Otte, L.J. van 't Veer, A. Berns, and M. van Lohuizen. 1997. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. Genes Dev. 11:226-40.
- Allemand, E., E. Batsche, and C. Muchardt. 2008. Splicing, transcription, and chromatin: a menage a trois. Curr Opin Genet Dev. 18:145-51.
- Ballestar, E., and M. Esteller. 2008. Epigenetic gene regulation in cancer. Adv Genet. 61:247-67.
- Banito, A., S.T. Rashid, J.C. Acosta, S. Li, C.F. Pereira, I. Geti, S. Pinho, J.C. Silva, V. Azuara, M. Walsh, L. Vallier, and J. Gil. 2009. Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev. 23:2134-9.

- Bannister, A.J., P. Zegerman, J.F. Partridge, E.A. Miska, J.O. Thomas, R.C. Allshire, and T. Kouzarides. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 410:120-4.
- Barradas, M., E. Anderton, J.C. Acosta, S. Li, A. Banito, M. Rodriguez-Niedenfuhr, G. Maertens, M. Banck, M.M. Zhou, M.J. Walsh, G. Peters, and J. Gil. 2009. Histone demethylase JMJD3 contributes to epigenetic control of INK4a/ARF by oncogenic RAS. Genes Dev. 23:1177-82.
- Barski, A., S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-resolution profiling of histone methylations in the human genome. Cell. 129:823-37.
- Bartkova, J., N. Rezaei, M. Liontos, P. Karakaidos, D. Kletsas, N. Issaeva, L.V. Vassiliou, E. Kolettas, K. Niforou, V.C. Zoumpourlis, M. Takaoka, H. Nakagawa, F. Tort, K. Fugger, F. Johansson, M. Sehested, C.L. Andersen, L. Dyrskjot, T. Orntoft, J. Lukas, C. Kittas, T. Helleday, T.D. Halazonetis, J. Bartek, and V.G. Gorgoulis. 2006. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature. 444:633-7.
- Bates, S., A.C. Phillips, P.A. Clark, F. Stott, G. Peters, R.L. Ludwig, and K.H. Vousden. 1998. p14ARF links the tumour suppressors RB and p53. Nature. 395:124-5.
- Bea, S., F. Tort, M. Pinyol, X. Puig, L. Hernandez, S. Hernandez, P.L. Fernandez, M. van Lohuizen, D. Colomer, and E. Campo. 2001. BMI-1 gene amplification and overexpression in hematological malignancies occur mainly in mantle cell lymphomas. Cancer Res. 61:2409-12.
- Beisel, C., A. Buness, I.M. Roustan-Espinosa, B. Koch, S. Schmitt, S.A. Haas, M. Hild, T. Katsuyama, and R. Paro. 2007. Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proc Natl Acad Sci U S A. 104:16615-20.
- Bernard, D., J.F. Martinez-Leal, S. Rizzo, D. Martinez, D. Hudson, T. Visakorpi, G. Peters, A. Carnero, D. Beach, and J. Gil. 2005. CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. Oncogene. 24:5543-51.
- Bernstein, B.E., T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, and E.S. Lander. 2006a. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 125:315-26.

- Bernstein, E., E.M. Duncan, O. Masui, J. Gil, E. Heard, and C.D. Allis. 2006b. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol. 26:2560-9.
- Besson, A., S.F. Dowdy, and J.M. Roberts. 2008. CDK inhibitors: cell cycle regulators and beyond. Dev Cell. 14:159-69.
- Blasco, M.A., H.W. Lee, M.P. Hande, E. Samper, P.M. Lansdorp, R.A. DePinho, and C.W. Greider. 1997. Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell. 91:25-34.
- Bloyer, S., G. Cavalli, H.W. Brock, and J.M. Dura. 2003. Identification and characterization of polyhomeotic PREs and TREs. Dev Biol. 261:426-42.
- Bodnar, A.G., M. Ouellette, M. Frolkis, S.E. Holt, C.P. Chiu, G.B. Morin, C.B. Harley, J.W. Shay, S. Lichtsteiner, and W.E. Wright. 1998. Extension of life-span by introduction of telomerase into normal human cells. Science. 279:349-52.
- Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, and R. Jaenisch. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 441:349-53.
- Bracken, A.P., N. Dietrich, D. Pasini, K.H. Hansen, and K. Helin. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 20:1123-36.
- Bracken, A.P., D. Kleine-Kohlbrecher, N. Dietrich, D. Pasini, G. Gargiulo, C. Beekman, K. Theilgaard-Monch, S. Minucci, B.T. Porse, J.C. Marine, K.H. Hansen, and K. Helin. 2007. The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev. 21:525-30.
- Bracken, A.P., D. Pasini, M. Capra, E. Prosperini, E. Colli, and K. Helin. 2003. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. Embo J. 22:5323-35.
- Braig, M., S. Lee, C. Loddenkemper, C. Rudolph, A.H. Peters, B. Schlegelberger, H. Stein, B. Dorken, T. Jenuwein, and C.A. Schmitt. 2005. Oncogene-induced senescence as an initial barrier in lymphoma development. Nature. 436:660-5.

- Breiling, A., B.M. Turner, M.E. Bianchi, and V. Orlando. 2001. General transcription factors bind promoters repressed by Polycomb group proteins. Nature. 412:651-5.
- Brookes, S., J. Rowe, A. Gutierrez Del Arroyo, J. Bond, and G. Peters. 2004. Contribution of p16(INK4a) to replicative senescence of human fibroblasts. Exp Cell Res. 298:549-59.
- Brookes, S., J. Rowe, M. Ruas, S. Llanos, P.A. Clark, M. Lomax, M.C. James, R. Vatcheva, S. Bates, K.H. Vousden, D. Parry, N. Gruis, N. Smit, W. Bergman, and G. Peters. 2002. INK4a-deficient human diploid fibroblasts are resistant to RAS-induced senescence. Embo J. 21:2936-45.
- Brown, C.J., A. Ballabio, J.L. Rupert, R.G. Lafreniere, M. Grompe, R. Tonlorenzi, and H.F. Willard. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature. 349:38-44.
- Brown, J.L., D.J. Grau, S.K. DeVido, and J.A. Kassis. 2005. An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Res. 33:5181-9.
- Brown, J.P., W. Wei, and J.M. Sedivy. 1997. Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. Science. 277:831-4.
- Bruggeman, S.W., D. Hulsman, E. Tanger, T. Buckle, M. Blom, J. Zevenhoven, O. van Tellingen, and M. van Lohuizen. 2007. Bmi1 controls tumor development in an Ink4a/Arf-independent manner in a mouse model for glioma. Cancer Cell. 12:328-41.
- Buchwald, G., P. van der Stoop, O. Weichenrieder, A. Perrakis, M. van Lohuizen, and T.K. Sixma. 2006. Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. Embo J. 25:2465-74.
- Cairns, B.R. 2009. The logic of chromatin architecture and remodelling at promoters. Nature. 461:193-8.
- Cales, C., M. Roman-Trufero, L. Pavon, I. Serrano, T. Melgar, M. Endoh, C. Perez, H. Koseki, and M. Vidal. 2008. Inactivation of the polycomb group protein Ring1B unveils an antiproliferative role in hematopoietic cell expansion and cooperation with tumorigenesis associated with Ink4a deletion. Mol Cell Biol. 28:1018-28.

- Campisi, J., and F. d'Adda di Fagagna. 2007. Cellular senescence: when bad things happen to good cells. Nat Rev Mol Cell Biol. 8:729-40.
- Cao, R., Y. Tsukada, and Y. Zhang. 2005. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol Cell. 20:845-54.
- Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R.S. Jones, and Y. Zhang. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science. 298:1039-43.
- Chamberlain, S.J., D. Yee, and T. Magnuson. 2008. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells. 26:1496-505.
- Chen, Z., L.C. Trotman, D. Shaffer, H.K. Lin, Z.A. Dotan, M. Niki, J.A. Koutcher, H.I. Scher, T. Ludwig, W. Gerald, C. Cordon-Cardo, and P.P. Pandolfi. 2005. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. Nature. 436:725-30.
- Cho, Y.W., T. Hong, S. Hong, H. Guo, H. Yu, D. Kim, T. Guszczynski, G.R. Dressler, T.D. Copeland, M. Kalkum, and K. Ge. 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J Biol Chem. 282:20395-406.
- Classon, M., and E. Harlow. 2002. The retinoblastoma tumour suppressor in development and cancer. Nat Rev Cancer. 2:910-7.
- Collado, M., J. Gil, A. Efeyan, C. Guerra, A.J. Schuhmacher, M. Barradas, A. Benguria, A. Zaballos, J.M. Flores, M. Barbacid, D. Beach, and M. Serrano. 2005. Tumour biology: senescence in premalignant tumours. Nature. 436:642.
- Core, L.J., and J.T. Lis. 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science. 319:1791-2.
- Core, N., S. Bel, S.J. Gaunt, M. Aurrand-Lions, J. Pearce, A. Fisher, and M. Djabali. 1997. Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development. 124:721-9.
- Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell. 111:185-96.
- d'Adda di Fagagna, F., P.M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. Von Zglinicki, G. Saretzki, N.P. Carter, and S.P. Jackson. 2003. A DNA damage checkpoint response in telomere-initiated senescence. Nature. 426:194-8.

- D'Alessio, J.A., K.J. Wright, and R. Tjian. 2009. Shifting players and paradigms in cell-specific transcription. Mol Cell. 36:924-31.
- Daujat, S., U. Zeissler, T. Waldmann, N. Happel, and R. Schneider. 2005. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. J Biol Chem. 280:38090-5.
- de la Mata, M., C.R. Alonso, S. Kadener, J.P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley, and A.R. Kornblihtt. 2003. A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell. 12:525-32.
- de Napoles, M., J.E. Mermoud, R. Wakao, Y.A. Tang, M. Endoh, R. Appanah, T.B. Nesterova, J. Silva, A.P. Otte, M. Vidal, H. Koseki, and N. Brockdorff. 2004. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell. 7:663-76.
- Deatrick, J., M. Daly, N.B. Randsholt, and H.W. Brock. 1991. The complex genetic locus polyhomeotic in Drosophila melanogaster potentially encodes two homologous zinc-finger proteins. Gene. 105:185-95.
- Dellino, G.I., Y.B. Schwartz, G. Farkas, D. McCabe, S.C. Elgin, and V. Pirrotta. 2004. Polycomb silencing blocks transcription initiation. Mol Cell. 13:887-93.
- Deng, Y., S.S. Chan, and S. Chang. 2008. Telomere dysfunction and tumour suppression: the senescence connection. Nat Rev Cancer. 8:450-8.
- Deshpande, A.M., J.D. Akunowicz, X.T. Reveles, B.B. Patel, E.A. Saria, R.G. Gorlick, S.L. Naylor, R.J. Leach, and M.F. Hansen. 2007. PHC3, a component of the hPRC-H complex, associates with E2F6 during G0 and is lost in osteosarcoma tumors. Oncogene. 26:1714-22.
- Dhawan, S., S.I. Tschen, and A. Bhushan. 2009. Bmi-1 regulates the Ink4a/Arf locus to control pancreatic beta-cell proliferation. Genes Dev. 23:906-11.
- Di Micco, R., M. Fumagalli, A. Cicalese, S. Piccinin, P. Gasparini, C. Luise, C. Schurra, M. Garre, P.G. Nuciforo, A. Bensimon, R. Maestro, P.G. Pelicci, and F. d'Adda di Fagagna. 2006. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature. 444:638-42.
- Dietrich, N., A.P. Bracken, E. Trinh, C.K. Schjerling, H. Koseki, J. Rappsilber, K. Helin, and K.H. Hansen. 2007. Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. Embo J. 26:1637-48.
- Dimri, G.P., X. Lee, G. Basile, M. Acosta, G. Scott, C. Roskelley, E.E. Medrano, M. Linskens, I. Rubelj, O. Pereira-Smith, and et al. 1995. A biomarker

that identifies senescent human cells in culture and in aging skin in vivo. Proc Natl Acad Sci U S A. 92:9363-7.

- Dimri, G.P., J.L. Martinez, J.J. Jacobs, P. Keblusek, K. Itahana, M. Van Lohuizen, J. Campisi, D.E. Wazer, and V. Band. 2002. The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. Cancer Res. 62:4736-45.
- Dura, J.M., H.W. Brock, and P. Santamaria. 1985. Polyhomeotic: a gene of Drosophila melanogaster required for correct expression of segmental identity. Mol Gen Genet. 198:213-20.
- Dura, J.M., N.B. Randsholt, J. Deatrick, I. Erk, P. Santamaria, J.D. Freeman, S.J. Freeman, D. Weddell, and H.W. Brock. 1987. A complex genetic locus, polyhomeotic, is required for segmental specification and epidermal development in D. melanogaster. Cell. 51:829-39.
- Dyson, N. 1998. The regulation of E2F by pRB-family proteins. Genes Dev. 12:2245-62.
- Ebert, A., G. Schotta, S. Lein, S. Kubicek, V. Krauss, T. Jenuwein, and G. Reuter. 2004. Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev. 18:2973-83.
- Elderkin, S., G.N. Maertens, M. Endoh, D.L. Mallery, N. Morrice, H. Koseki, G. Peters, N. Brockdorff, and K. Hiom. 2007. A phosphorylated form of Mel-18 targets the Ring1B histone H2A ubiquitin ligase to chromatin. Mol Cell. 28:107-20.
- Endoh, M., T.A. Endo, T. Endoh, Y. Fujimura, O. Ohara, T. Toyoda, A.P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki. 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development. 135:1513-24.
- Ezhkova, E., H.A. Pasolli, J.S. Parker, N. Stokes, I.H. Su, G. Hannon, A. Tarakhovsky, and E. Fuchs. 2009. Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell. 136:1122-35.
- Faast, R., J. White, P. Cartwright, L. Crocker, B. Sarcevic, and S. Dalton. 2004. Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). Oncogene. 23:491-502.
- Fauvarque, M.O., V. Zuber, and J.M. Dura. 1995. Regulation of polyhomeotic transcription may involve local changes in chromatin activity in Drosophila. Mech Dev. 52:343-55.

- Fernandez, P.C., S.R. Frank, L. Wang, M. Schroeder, S. Liu, J. Greene, A. Cocito, and B. Amati. 2003. Genomic targets of the human c-Myc protein. Genes Dev. 17:1115-29.
- Fernandez-Capetillo, O., and M. Murga. 2008. Why cells respond differently to DNA damage: a chromatin perspective. Cell Cycle. 7:980-3.
- Fischle, W., Y. Wang, S.A. Jacobs, Y. Kim, C.D. Allis, and S. Khorasanizadeh. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. 17:1870-81.
- Francis, N.J., R.E. Kingston, and C.L. Woodcock. 2004. Chromatin compaction by a polycomb group protein complex. Science. 306:1574-7.
- Francis, N.J., A.J. Saurin, Z. Shao, and R.E. Kingston. 2001. Reconstitution of a functional core polycomb repressive complex. Mol Cell. 8:545-56.
- Gaestel, M. 2006. MAPKAP kinases - MKs - two's company, three's a crowd. Nat Rev Mol Cell Biol. 7:120-30.
- Gearhart, M.D., C.M. Corcoran, J.A. Wamstad, and V.J. Bardwell. 2006. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol Cell Biol. 26:6880-9.
- Gil, J., D. Bernard, D. Martinez, and D. Beach. 2004. Polycomb CBX7 has a unifying role in cellular lifespan. Nat Cell Biol. 6:67-72.
- Gil, J., D. Bernard, and G. Peters. 2005. Role of polycomb group proteins in stem cell self-renewal and cancer. DNA Cell Biol. 24:117-25.
- Gil, J., and G. Peters. 2006. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nat Rev Mol Cell Biol. 7:667-77.
- Goldberg, A.D., C.D. Allis, and E. Bernstein. 2007. Epigenetics: a landscape takes shape. Cell. 128:635-8.
- Gonzalez, S., D.G. Pisano, and M. Serrano. 2008. Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. Cell Cycle. 7:2601-8.
- Grimaud, C., F. Bantignies, M. Pal-Bhadra, P. Ghana, U. Bhadra, and G. Cavalli. 2006. RNAi components are required for nuclear clustering of Polycomb group response elements. Cell. 124:957-71.
- Gunster, M.J., D.P. Satijn, K.M. Hamer, J.L. den Blaauwen, D. de Bruijn, M.J. Alkema, M. van Lohuizen, R. van Driel, and A.P. Otte. 1997. Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. Mol Cell Biol. 17:2326-35.

- Gunther, M., M. Laithier, and O. Brison. 2000. A set of proteins interacting with transcription factor Sp1 identified in a two-hybrid screening. Mol Cell Biochem. 210:131-42.
- Guo, W.J., S. Datta, V. Band, and G.P. Dimri. 2007a. Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. Mol Biol Cell. 18:536-46.
- Guo, W.J., M.S. Zeng, A. Yadav, L.B. Song, B.H. Guo, V. Band, and G.P. Dimri. 2007b. Mel-18 acts as a tumor suppressor by repressing Bmi-1 expression and down-regulating Akt activity in breast cancer cells. Cancer Res. 67:5083-9.
- Gupta, P.B., C.L. Chaffer, and R.A. Weinberg. 2009. Cancer stem cells: mirage or reality? Nat Med. 15:1010-2.
- Guttman, M., I. Amit, M. Garber, C. French, M.F. Lin, D. Feldser, M. Huarte, O. Zuk, B.W. Carey, J.P. Cassady, M.N. Cabili, R. Jaenisch, T.S. Mikkelsen, T. Jacks, N. Hacohen, B.E. Bernstein, M. Kellis, A. Regev, J.L. Rinn, and E.S. Lander. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 458:223-7.
- Hager, G.L., J.G. McNally, and T. Misteli. 2009. Transcription dynamics. Mol Cell. 35:741-53.
- Haupt, Y., W.S. Alexander, G. Barri, S.P. Klinken, and J.M. Adams. 1991. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. Cell. 65:753-63.
- Hayflick, L., and P.S. Moorhead. 1961. The serial cultivation of human diploid cell strains. Exp Cell Res. 25:585-621.
- He, S., T. Iwashita, J. Buchstaller, A.V. Molofsky, D. Thomas, and S.J. Morrison. 2009. Bmi-1 over-expression in neural stem/progenitor cells increases proliferation and neurogenesis in culture but has little effect on these functions in vivo. Dev Biol. 328:257-72.
- Helgadottir, A., G. Thorleifsson, A. Manolescu, S. Gretarsdottir, T. Blondal, A. Jonasdottir, A. Jonasdottir, A. Sigurdsson, A. Baker, A. Palsson, G. Masson, D.F. Gudbjartsson, K.P. Magnusson, K. Andersen, A.I. Levey, V.M. Backman, S. Matthiasdottir, T. Jonsdottir, S. Palsson, H. Einarsdottir, S. Gunnarsdottir, A. Gylfason, V. Vaccarino, W.C. Hooper, M.P. Reilly, C.B. Granger, H. Austin, D.J. Rader, S.H. Shah, A.A. Quyyumi, J.R. Gulcher, G. Thorgeirsson, U. Thorsteinsdottir, A. Kong, and K. Stefansson. 2007. A common variant on

chromosome 9p21 affects the risk of myocardial infarction. Science. 316:1491-3.

- Hernandez-Munoz, I., P. Taghavi, C. Kuijl, J. Neefjes, and M. van Lohuizen. 2005. Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. Mol Cell Biol. 25:11047-58.
- Hodges, C., L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. 2009. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science. 325:626-8.
- Isono, K., Y. Fujimura, J. Shinga, M. Yamaki, O.W. J, Y. Takihara, Y. Murahashi, Y. Takada, Y. Mizutani-Koseki, and H. Koseki. 2005a. Mammalian polyhomeotic homologues Phc2 and Phc1 act in synergy to mediate polycomb repression of Hox genes. Mol Cell Biol. 25:6694-706.
- Isono, K., Y. Mizutani-Koseki, T. Komori, M.S. Schmidt-Zachmann, and H. Koseki. 2005b. Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev. 19:536-41.
- Issaeva, I., Y. Zonis, T. Rozovskaia, K. Orlovsky, C.M. Croce, T. Nakamura, A. Mazo, L. Eisenbach, and E. Canaani. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol Cell Biol. 27:1889-903.
- Iwama, A., H. Oguro, M. Negishi, Y. Kato, Y. Morita, H. Tsukui, H. Ema, T. Kamijo, Y. Katoh-Fukui, H. Koseki, M. van Lohuizen, and H. Nakauchi. 2004. Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. Immunity. 21:843-51.
- Jacobs, J.J., K. Kieboom, S. Marino, R.A. DePinho, and M. van Lohuizen. 1999a. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. Nature. 397:164-8.
- Jacobs, J.J., B. Scheijen, J.W. Voncken, K. Kieboom, A. Berns, and M. van Lohuizen. 1999b. Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. Genes Dev. 13:2678-90.
- Jaenisch, R., and R. Young. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell. 132:567-82.
- Janzen, V., R. Forkert, H.E. Fleming, Y. Saito, M.T. Waring, D.M. Dombkowski, T. Cheng, R.A. DePinho, N.E. Sharpless, and D.T. Scadden.

2006. Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a. Nature. 443:421-6.

- Johnson, K.M., J. Wang, A. Smallwood, C. Arayata, and M. Carey. 2002. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. Genes Dev. 16:1852-63.
- Kajiume, T., Y. Ninomiya, H. Ishihara, R. Kanno, and M. Kanno. 2004. Polycomb group gene mel-18 modulates the self-renewal activity and cell cycle status of hematopoietic stem cells. Exp Hematol. 32:571-8.
- Kalantry, S., S. Purushothaman, R.B. Bowen, J. Starmer, and T. Magnuson. 2009. Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature. 460:647-51.
- Kamb, A., N.A. Gruis, J. Weaver-Feldhaus, Q. Liu, K. Harshman, S.V. Tavtigian, E. Stockert, R.S. Day, 3rd, B.E. Johnson, and M.H. Skolnick. 1994. A cell cycle regulator potentially involved in genesis of many tumor types. Science. 264:436-40.
- Kamijo, T., F. Zindy, M.F. Roussel, D.E. Quelle, J.R. Downing, R.A. Ashmun, G. Grosveld, and C.J. Sherr. 1997. Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. Cell. 91:649-59.
- Kang, M.K., R.H. Kim, S.J. Kim, F.K. Yip, K.H. Shin, G.P. Dimri, R. Christensen, T. Han, and N.H. Park. 2007. Elevated Bmi-1 expression is associated with dysplastic cell transformation during oral carcinogenesis and is required for cancer cell replication and survival. Br J Cancer. 96:126-33.
- Kennedy, C.H., H.I. Pass, and J.B. Mitchell. 2003. Expression of human MutT homologue (hMTH1) protein in primary non-small-cell lung carcinomas and histologically normal surrounding tissue. Free Radic Biol Med. 34:1447-57.
- Kennison, J.A., and J.W. Tamkun. 1988. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc Natl Acad Sci U S A. 85:8136-40.
- Khalil, A.M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B.E. Bernstein, A. van Oudenaarden, A. Regev, E.S. Lander, and J.L. Rinn. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A.

- Kia, S.K., M.M. Gorski, S. Giannakopoulos, and C.P. Verrijzer. 2008. SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. Mol Cell Biol. 28:3457-64.
- Kim, C.A., M. Gingery, R.M. Pilpa, and J.U. Bowie. 2002. The SAM domain of polyhomeotic forms a helical polymer. Nat Struct Biol. 9:453-7.
- Kim, C.A., M.R. Sawaya, D. Cascio, W. Kim, and J.U. Bowie. 2005. Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. J Biol Chem. 280:27769-75.
- Kim, D.H., L.M. Villeneuve, K.V. Morris, and J.J. Rossi. 2006. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol. 13:793-7.
- Kim, W.Y., and N.E. Sharpless. 2006. The regulation of INK4/ARF in cancer and aging. Cell. 127:265-75.
- Kirmizis, A., S.M. Bartley, and P.J. Farnham. 2003. Identification of the polycomb group protein SU(Z)12 as a potential molecular target for human cancer therapy. Mol Cancer Ther. 2:113-21.
- Kleer, C.G., Q. Cao, S. Varambally, R. Shen, I. Ota, S.A. Tomlins, D. Ghosh, R.G. Sewalt, A.P. Otte, D.F. Hayes, M.S. Sabel, D. Livant, S.J. Weiss, M.A. Rubin, and A.M. Chinnaiyan. 2003. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci U S A. 100:11606-11.
- Kotake, Y., Y. Zeng, and Y. Xiong. 2009. DDB1-CUL4 and MLL1 mediate oncogene-induced p16INK4a activation. Cancer Res. 69:1809-14.
- Krimpenfort, P., A. Ijpenberg, J.Y. Song, M. van der Valk, M. Nawijn, J. Zevenhoven, and A. Berns. 2007. p15Ink4b is a critical tumour suppressor in the absence of p16Ink4a. Nature. 448:943-6.
- Krimpenfort, P., K.C. Quon, W.J. Mooi, A. Loonstra, and A. Berns. 2001. Loss of p16Ink4a confers susceptibility to metastatic melanoma in mice. Nature. 413:83-6.
- Krishnamurthy, J., M.R. Ramsey, K.L. Ligon, C. Torrice, A. Koh, S. Bonner-Weir, and N.E. Sharpless. 2006. p16INK4a induces an age-dependent decline in islet regenerative potential. Nature. 443:453-7.
- Krishnamurthy, J., C. Torrice, M.R. Ramsey, G.I. Kovalev, K. Al-Regaiey, L. Su, and N.E. Sharpless. 2004. Ink4a/Arf expression is a biomarker of aging. J Clin Invest. 114:1299-307.

- Krumlauf, R. 1994. Hox genes in vertebrate development. Cell. 78:191-201.
- Kuilman, T., C. Michaloglou, L.C. Vredeveld, S. Douma, R. van Doorn, C.J. Desmet, L.A. Aarden, W.J. Mooi, and D.S. Peeper. 2008. Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. Cell. 133:1019-31.
- Kuilman, T., and D.S. Peeper. 2009. Senescence-messaging secretome: SMS-ing cellular stress. Nat Rev Cancer. 9:81-94.
- Kuzmichev, A., T. Jenuwein, P. Tempst, and D. Reinberg. 2004. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Mol Cell. 14:183-93.
- Kuzmichev, A., R. Margueron, A. Vaquero, T.S. Preissner, M. Scher, A. Kirmizis, X. Ouyang, N. Brockdorff, C. Abate-Shen, P. Farnham, and D. Reinberg. 2005. Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. Proc Natl Acad Sci U S A. 102:1859-64.
- Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 16:2893-905.
- Kyba, M., and H.W. Brock. 1998a. The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. Mol Cell Biol. 18:2712-20.
- Kyba, M., and H.W. Brock. 1998b. The SAM domain of polyhomeotic, RAE28, and scm mediates specific interactions through conserved residues. Dev Genet. 22:74-84.
- Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 410:116-20.
- Lagarou, A., A. Mohd-Sarip, Y.M. Moshkin, G.E. Chalkley, K. Bezstarosti, J.A. Demmers, and C.P. Verrijzer. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. 22:2799-810.
- Lan, F., P.E. Bayliss, J.L. Rinn, J.R. Whetstine, J.K. Wang, S. Chen, S. Iwase, R. Alpatov, I. Issaeva, E. Canaani, T.M. Roberts, H.Y. Chang, and Y. Shi. 2007.

A histone H3 lysine 27 demethylase regulates animal posterior development. Nature. 449:689-94.

- Lanzuolo, C., V. Roure, J. Dekker, F. Bantignies, and V. Orlando. 2007. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol. 9:1167-74.
- Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M.A. Caligiuri, and J.E. Dick. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 367:645-8.
- Lechtenberg, B.C., M.D. Allen, T.J. Rutherford, S.M. Freund, and M. Bycroft. 2009. Solution structure of the FCS zinc finger domain of the human polycomb group protein L(3)mbt-like 2. Protein Sci. 18:657-61.
- Lee, M.G., R. Villa, P. Trojer, J. Norman, K.P. Yan, D. Reinberg, L. Di Croce, and R. Shiekhattar. 2007. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science. 318:447-50.
- Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker, B. Yuan, G.W. Bell, E. Herbolsheimer, N.M. Hannett, K. Sun, D.T. Odom, A.P. Otte, T.L. Volkert, D.P. Bartel, D.A. Melton, D.K. Gifford, R. Jaenisch, and R.A. Young. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 125:301-13.
- Leeb, M., and A. Wutz. 2007. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol. 178:219-29.
- Lessard, J., and G. Sauvageau. 2003. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature. 423:255-60.
- Lessard, J., A. Schumacher, U. Thorsteinsdottir, M. van Lohuizen, T. Magnuson, and G. Sauvageau. 1999. Functional antagonism of the Polycomb-Group genes eed and Bmi1 in hemopoietic cell proliferation. Genes Dev. 13:2691-703.
- Leung, C., M. Lingbeek, O. Shakhova, J. Liu, E. Tanger, P. Saremaslani, M. Van Lohuizen, and S. Marino. 2004. Bmi1 is essential for cerebellar development and is overexpressed in human medulloblastomas. Nature. 428:337-41.

- Levine, S.S., A. Weiss, H. Erdjument-Bromage, Z. Shao, P. Tempst, and R.E. Kingston. 2002. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol Cell Biol. 22:6070-8.
- Lewis, E.B. 1978. A gene complex controlling segmentation in Drosophila. Nature. 276:565-70.
- Lewis, J.L., W. Chinswangwatanakul, B. Zheng, S.B. Marley, D.X. Nguyen, N.C. Cross, L. Banerji, J. Glassford, N.S. Thomas, J.M. Goldman, E.W. Lam, and M.Y. Gordon. 2001. The influence of INK4 proteins on growth and self-renewal kinetics of hematopoietic progenitor cells. Blood. 97:2604-10.
- Li, B., J. Zhou, P. Liu, J. Hu, H. Jin, Y. Shimono, M. Takahashi, and G. Xu. 2007. Polycomb protein Cbx4 promotes SUMO modification of de novo DNA methyltransferase Dnmt3a. Biochem J. 405:369-78.
- Li, H., M. Collado, A. Villasante, K. Strati, S. Ortega, M. Canamero, M.A. Blasco, and M. Serrano. 2009. The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature. 460:1136-9.
- Li, T., J.F. Hu, X. Qiu, J. Ling, H. Chen, S. Wang, A. Hou, T.H. Vu, and A.R. Hoffman. 2008. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol. 28:6473-82.
- Li, Z., R. Cao, M. Wang, M.P. Myers, Y. Zhang, and R.M. Xu. 2006. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. J Biol Chem. 281:20643-9.
- Lin, A.W., M. Barradas, J.C. Stone, L. van Aelst, M. Serrano, and S.W. Lowe. 1998. Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. Genes Dev. 12:3008-19.
- Listerman, I., A.K. Sapra, and K.M. Neugebauer. 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat Struct Mol Biol. 13:815-22.
- Liu, J., L. Cao, J. Chen, S. Song, I.H. Lee, C. Quijano, H. Liu, K. Keyvanfar, H. Chen, L.Y. Cao, B.H. Ahn, N.G. Kumar, Rovira, II, X.L. Xu, M. van Lohuizen, N. Motoyama, C.X. Deng, and T. Finkel. 2009. Bmi1 regulates mitochondrial function and the DNA damage response pathway. Nature. 459:387-92.
- Liu, L., L.G. Andrews, and T.O. Tollefsbol. 2006. Loss of the human polycomb group protein BMI1 promotes cancer-specific cell death. Oncogene. 25:4370-5.

- Lu, T., and T. Finkel. 2008. Free radicals and senescence. Exp Cell Res. 314:1918-22.
- Maddika, S., and J. Chen. 2009. Protein kinase DYRK2 is a scaffold that facilitates assembly of an E3 ligase. Nat Cell Biol. 11:409-19.
- Maertens, G.N., S. El Messaoudi-Aubert, T. Racek, J.K. Stock, J. Nicholls, M. Rodriguez-Niedenfuhr, J. Gil, and G. Peters. 2009. Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. PLoS One. 4:e6380.
- Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil, C.L. Woodcock, B.D. Dynlacht, and D. Reinberg. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell. 32:503-18.
- Marion, R.M., K. Strati, H. Li, M. Murga, R. Blanco, S. Ortega, O. Fernandez-Capetillo, M. Serrano, and M.A. Blasco. 2009. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature. 460:1149-53.
- Martianov, I., A. Ramadass, A. Serra Barros, N. Chow, and A. Akoulitchev. 2007. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 445:666-70.
- Martinez, A.M., B. Schuettengruber, S. Sakr, A. Janic, C. Gonzalez, and G. Cavalli. 2009. Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling. Nat Genet.
- Masutomi, K., E.Y. Yu, S. Khurts, I. Ben-Porath, J.L. Currier, G.B. Metz, M.W. Brooks, S. Kaneko, S. Murakami, J.A. DeCaprio, R.A. Weinberg, S.A. Stewart, and W.C. Hahn. 2003. Telomerase maintains telomere structure in normal human cells. Cell. 114:241-53.
- Matheu, A., C. Pantoja, A. Efeyan, L.M. Criado, J. Martin-Caballero, J.M. Flores, P. Klatt, and M. Serrano. 2004. Increased gene dosage of Ink4a/Arf results in cancer resistance and normal aging. Genes Dev. 18:2736-46.
- Matsuoka, S., B.A. Ballif, A. Smogorzewska, E.R. McDonald, 3rd, K.E. Hurov, J. Luo, C.E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S.P. Gygi, and S.J. Elledge. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science. 316:1160-6.
- McPherson, R., A. Pertsemlidis, N. Kavaslar, A. Stewart, R. Roberts, D.R. Cox, D.A. Hinds, L.A. Pennacchio, A. Tybjaerg-Hansen, A.R. Folsom, E.

Boerwinkle, H.H. Hobbs, and J.C. Cohen. 2007. A common allele on chromosome 9 associated with coronary heart disease. Science. 316:1488-91.

- Michaloglou, C., L.C. Vredeveld, M.S. Soengas, C. Denoyelle, T. Kuilman, C.M. van der Horst, D.M. Majoor, J.W. Shay, W.J. Mooi, and D.S. Peeper. 2005. BRAFE600-associated senescence-like cell cycle arrest of human naevi. Nature. 436:720-4.
- Miki, J., Y. Fujimura, H. Koseki, and T. Kamijo. 2007. Polycomb complexes regulate cellular senescence by repression of ARF in cooperation with E2F3. Genes Cells. 12:1371-82.
- Mikkelsen, T.S., J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B.E. Bernstein, R. Jaenisch, E.S. Lander, and A. Meissner. 2008. Dissecting direct reprogramming through integrative genomic analysis. Nature. 454:49-55.
- Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, and B.E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 448:553-60.
- Min, J., Y. Zhang, and R.M. Xu. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. 17:1823-8.
- Mo, J.Y., H. Maki, and M. Sekiguchi. 1992. Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. Proc Natl Acad Sci U S A. 89:11021-5.
- Mollaaghababa, R., L. Sipos, S.Y. Tiong, O. Papoulas, J.A. Armstrong, J.W. Tamkun, and W. Bender. 2001. Mutations in Drosophila heat shock cognate 4 are enhancers of Polycomb. Proc Natl Acad Sci U S A. 98:3958-63.
- Molofsky, A.V., R. Pardal, T. Iwashita, I.K. Park, M.F. Clarke, and S.J. Morrison. 2003. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature. 425:962-7.
- Molofsky, A.V., S.G. Slutsky, N.M. Joseph, S. He, R. Pardal, J. Krishnamurthy, N.E. Sharpless, and S.J. Morrison. 2006. Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. Nature. 443:448-52.

- Montgomery, N.D., D. Yee, A. Chen, S. Kalantry, S.J. Chamberlain, A.P. Otte, and T. Magnuson. 2005. The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr Biol. 15:942-7.
- Muller, J., C.M. Hart, N.J. Francis, M.L. Vargas, A. Sengupta, B. Wild, E.L. Miller, M.B. O'Connor, R.E. Kingston, and J.A. Simon. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell. 111:197-208.
- Murga, M., I. Jaco, Y. Fan, R. Soria, B. Martinez-Pastor, M. Cuadrado, S.M. Yang, M.A. Blasco, A.I. Skoultchi, and O. Fernandez-Capetillo. 2007. Global chromatin compaction limits the strength of the DNA damage response. J Cell Biol. 178:1101-8.
- Myohanen, S., and S.B. Baylin. 2001. Sequence-specific DNA binding activity of RNA helicase A to the p16INK4a promoter. J Biol Chem. 276:1634-42.
- Nakabeppu, Y., K. Kajitani, K. Sakamoto, H. Yamaguchi, and D. Tsuchimoto. 2006. MTH1, an oxidized purine nucleoside triphosphatase, prevents the cytotoxicity and neurotoxicity of oxidized purine nucleotides. DNA Repair (Amst). 5:761-72.
- Nekrasov, M., B. Wild, and J. Muller. 2005. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep. 6:348-53.
- Ng, J., C.M. Hart, K. Morgan, and J.A. Simon. 2000. A Drosophila ESC-E(Z) protein complex is distinct from other polycomb group complexes and contains covalently modified ESC. Mol Cell Biol. 20:3069-78.
- Nguyen, N., X. Zhang, N. Olashaw, and E. Seto. 2004. Molecular cloning and functional characterization of the transcription factor YY2. J Biol Chem. 279:25927-34.
- Nicholls, J. 2006. Investigating protein complexes that are involved in the function and regulation of the human INK4a/ARF locus. PhD Thesis, University College London, UK.
- Nomura, M., Y. Takihara, and K. Shimada. 1994. Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: one of the early inducible clones encodes a novel protein sharing several highly homologous regions with a Drosophila polyhomeotic protein. Differentiation. 57:39-50.
- O'Carroll, D., S. Erhardt, M. Pagani, S.C. Barton, M.A. Surani, and T. Jenuwein. 2001. The polycomb-group gene Ezh2 is required for early mouse development. Mol Cell Biol. 21:4330-6.

- Ogawa, H., K. Ishiguro, S. Gaubatz, D.M. Livingston, and Y. Nakatani. 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science. 296:1132-6.
- Ohno, K., D. McCabe, B. Czermin, A. Imhof, and V. Pirrotta. 2008. ESC, ESCL and their roles in Polycomb Group mechanisms. Mech Dev. 125:527-41.
- Ohta, H., A. Sawada, J.Y. Kim, S. Tokimasa, S. Nishiguchi, R.K. Humphries, J. Hara, and Y. Takihara. 2002. Polycomb group gene rae28 is required for sustaining activity of hematopoietic stem cells. J Exp Med. 195:759-70.
- Orford, K.W., and D.T. Scadden. 2008. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nat Rev Genet. 9:115-28.
- Pandey, R.R., T. Mondal, F. Mohammad, S. Enroth, L. Redrup, J. Komorowski, T. Nagano, D. Mancini-Dinardo, and C. Kanduri. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 32:232-46.
- Pandya-Jones, A., and D.L. Black. 2009. Co-transcriptional splicing of constitutive and alternative exons. Rna. 15:1896-908.
- Papp, B., and J. Muller. 2006. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20:2041-54.
- Park, I.K., D. Qian, M. Kiel, M.W. Becker, M. Pihalja, I.L. Weissman, S.J. Morrison, and M.F. Clarke. 2003. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature. 423:302-5.
- Parrinello, S., E. Samper, A. Krtolica, J. Goldstein, S. Melov, and J. Campisi. 2003. Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. Nat Cell Biol. 5:741-7.
- Pasini, D., A.P. Bracken, J.B. Hansen, M. Capillo, and K. Helin. 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol. 27:3769-79.
- Pasini, D., A.P. Bracken, M.R. Jensen, E. Lazzerini Denchi, and K. Helin. 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. Embo J. 23:4061-71.
- Pasmant, E., I. Laurendeau, D. Heron, M. Vidaud, D. Vidaud, and I. Bieche. 2007. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an

antisense noncoding RNA whose expression coclusters with ARF. Cancer Res. 67:3963-9.

- Pear, W.S., G.P. Nolan, M.L. Scott, and D. Baltimore. 1993. Production of high-titer helper-free retroviruses by transient transfection. Proc Natl Acad Sci U S A. 90:8392-6.
- Peters, G. 2008. Tumor suppression for ARFicionados: the relative contributions of p16INK4a and p14ARF in melanoma. J Natl Cancer Inst. 100:757-9.
- Peterson, A.J., M. Kyba, D. Bornemann, K. Morgan, H.W. Brock, and J. Simon. 1997. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. Mol Cell Biol. 17:6683-92.
- Plath, K., D. Talbot, K.M. Hamer, A.P. Otte, T.P. Yang, R. Jaenisch, and B. Panning. 2004. Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. J Cell Biol. 167:1025-35.
- Poux, S., R. Melfi, and V. Pirrotta. 2001. Establishment of Polycomb silencing requires a transient interaction between PC and ESC. Genes Dev. 15:2509-14.
- Quelle, D.E., F. Zindy, R.A. Ashmun, and C.J. Sherr. 1995. Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell. 83:993-1000.
- Rai, P., T.T. Onder, J.J. Young, J.L. McFaline, B. Pang, P.C. Dedon, and R.A. Weinberg. 2009. Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence. Proc Natl Acad Sci U S A. 106:169-74.
- Ramirez, R.D., C.P. Morales, B.S. Herbert, J.M. Rohde, C. Passons, J.W. Shay, and W.E. Wright. 2001. Putative telomere-independent mechanisms of replicative aging reflect inadequate growth conditions. Genes Dev. 15:398-403.
- Rea, S., F. Eisenhaber, D. O'Carroll, B.D. Strahl, Z.W. Sun, M. Schmid, S. Opravil, K. Mechtler, C.P. Ponting, C.D. Allis, and T. Jenuwein. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 406:593-9.
- Ressler, S., J. Bartkova, H. Niederegger, J. Bartek, K. Scharffetter-Kochanek, P. Jansen-Durr, and M. Wlaschek. 2006. p16INK4A is a robust in vivo biomarker of cellular aging in human skin. Aging Cell. 5:379-89.
- Reya, T., S.J. Morrison, M.F. Clarke, and I.L. Weissman. 2001. Stem cells, cancer, and cancer stem cells. Nature. 414:105-11.

- Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 17:1030-2.
- Ringrose, L. 2007. Polycomb comes of age: genome-wide profiling of target sites. Curr Opin Cell Biol. 19:290-7.
- Ringrose, L., H. Ehret, and R. Paro. 2004. Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol Cell. 16:641-53.
- Rinn, J.L., M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, and H.Y. Chang. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 129:1311-23.
- Rodier, F., J.P. Coppe, C.K. Patil, W.A. Hoeijmakers, D.P. Munoz, S.R. Raza, A. Freund, E. Campeau, A.R. Davalos, and J. Campisi. 2009. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. Nat Cell Biol. 11:973-9.
- Rohozinski, J., and C.E. Bishop. 2004. The mouse juvenile spermatogonial depletion (jsd) phenotype is due to a mutation in the X-derived retrogene, mUtp14b. Proc Natl Acad Sci U S A. 101:11695-700.
- Roy, N., A. Van Eynde, L. Beke, M. Nuytten, and M. Bollen. 2007. The transcriptional repression by NIPP1 is mediated by Polycomb group proteins. Biochim Biophys Acta. 1769:541-5.
- Ruas, M., and G. Peters. 1998. The p16INK4a/CDKN2A tumor suppressor and its relatives. Biochim Biophys Acta. 1378:F115-77.
- Sakamoto, Y., S. Watanabe, T. Ichimura, M. Kawasuji, H. Koseki, H. Baba, and M. Nakao. 2007. Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation. J Biol Chem. 282:16391-400.
- Sancho, M., E. Diani, M. Beato, and A. Jordan. 2008. Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet. 4:e1000227.
- Sandler, L., D.L. Lindsley, B. Nicoletti, and G. Trippa. 1968. Mutants affecting meiosis in natural populations of Drosophila melanogaster. Genetics. 60:525-58.

- Sasaki, M., J. Yamaguchi, K. Itatsu, H. Ikeda, and Y. Nakanuma. 2008. Over-expression of polycomb group protein EZH2 relates to decreased expression of p16 INK4a in cholangiocarcinogenesis in hepatolithiasis. J Pathol. 215:175-83.
- Satijn, D.P., M.J. Gunster, J. van der Vlag, K.M. Hamer, W. Schul, M.J. Alkema, A.J. Saurin, P.S. Freemont, R. van Driel, and A.P. Otte. 1997. RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. Mol Cell Biol. 17:4105-13.
- Satijn, D.P., K.M. Hamer, J. den Blaauwen, and A.P. Otte. 2001. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. Mol Cell Biol. 21:1360-9.
- Saurin, A.J., Z. Shao, H. Erdjument-Bromage, P. Tempst, and R.E. Kingston. 2001. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature. 412:655-60.
- Saurin, A.J., C. Shiels, J. Williamson, D.P. Satijn, A.P. Otte, D. Sheer, and P.S. Freemont. 1998. The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. J Cell Biol. 142:887-98.
- Saxena, R., B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly, T.E. Hughes, L. Groop, D. Altshuler, P. Almgren, J.C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H.N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E.K. Speliotes, M.R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S.B. Gabriel, G.W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke, and S. Purcell. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 316:1331-6.
- Schlesinger, Y., R. Straussman, I. Keshet, S. Farkash, M. Hecht, J. Zimmerman, E. Eden, Z. Yakhini, E. Ben-Shushan, B.E. Reubinoff, Y. Bergman, I. Simon, and H. Cedar. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet. 39:232-6.

- Schmitt, S., M. Prestel, and R. Paro. 2005. Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. 19:697-708.
- Schoeftner, S., A.K. Sengupta, S. Kubicek, K. Mechtler, L. Spahn, H. Koseki, T. Jenuwein, and A. Wutz. 2006. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. Embo J. 25:3110-22.
- Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, and T. Jenuwein. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. 18:1251-62.
- Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc, and G. Cavalli. 2007. Genome regulation by polycomb and trithorax proteins. Cell. 128:735-45.
- Schumacher, A., C. Faust, and T. Magnuson. 1996. Positional cloning of a global regulator of anterior-posterior patterning in mice. Nature. 384:648.
- Schwartz, S., E. Meshorer, and G. Ast. 2009. Chromatin organization marks exon-intron structure. Nat Struct Mol Biol. 16:990-5.
- Schwartz, Y.B., T.G. Kahn, D.A. Nix, X.Y. Li, R. Bourgon, M. Biggin, and V. Pirrotta. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet. 38:700-5.
- Schwartz, Y.B., and V. Pirrotta. 2007. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet. 8:9-22.
- Schwartz, Y.B., and V. Pirrotta. 2008. Polycomb complexes and epigenetic states. Curr Opin Cell Biol. 20:266-73.
- Schwermann, J., C. Rathinam, M. Schubert, S. Schumacher, F. Noyan, H. Koseki, A. Kotlyarov, C. Klein, and M. Gaestel. 2009. MAPKAP kinase MK2 maintains self-renewal capacity of haematopoietic stem cells. Embo J. 28:1392-406.
- Scott, C.L., J. Gil, E. Hernando, J. Teruya-Feldstein, M. Narita, D. Martinez, T. Visakorpi, D. Mu, C. Cordon-Cardo, G. Peters, D. Beach, and S.W. Lowe. 2007a. Role of the chromobox protein CBX7 in lymphomagenesis. Proc Natl Acad Sci U S A. 104:5389-94.
- Scott, L.J., K.L. Mohlke, L.L. Bonnycastle, C.J. Willer, Y. Li, W.L. Duren, M.R. Erdos, H.M. Stringham, P.S. Chines, A.U. Jackson, L. Prokunina-Olsson, C.J. Ding, A.J. Swift, N. Narisu, T. Hu, R. Pruim, R. Xiao, X.Y. Li, K.N. Conneely, N.L. Riebow, A.G. Sprau, M. Tong, P.P. White, K.N. Hetrick, M.W. Barnhart, C.W. Bark, J.L. Goldstein, L. Watkins, F. Xiang, J. Saramies, T.A.

Buchanan, R.M. Watanabe, T.T. Valle, L. Kinnunen, G.R. Abecasis, E.W. Pugh, K.F. Doheny, R.N. Bergman, J. Tuomilehto, F.S. Collins, and M. Boehnke. 2007b. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 316:1341-5.

- Serrano, M., A.W. Lin, M.E. McCurrach, D. Beach, and S.W. Lowe. 1997. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell. 88:593-602.
- Sessa, L., A. Breiling, G. Lavorgna, L. Silvestri, G. Casari, and V. Orlando. 2007. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. Rna. 13:223-39.
- Sewalt, R.G., J. van der Vlag, M.J. Gunster, K.M. Hamer, J.L. den Blaauwen, D.P. Satijn, T. Hendrix, R. van Driel, and A.P. Otte. 1998. Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. Mol Cell Biol. 18:3586-95.
- Shao, Z., F. Raible, R. Mollaaghababa, J.R. Guyon, C.T. Wu, W. Bender, and R.E. Kingston. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell. 98:37-46.
- Sharpless, N.E. 2005. INK4a/ARF: a multifunctional tumor suppressor locus. Mutat Res. 576:22-38.
- Sharpless, N.E., N. Bardeesy, K.H. Lee, D. Carrasco, D.H. Castrillon, A.J. Aguirre, E.A. Wu, J.W. Horner, and R.A. DePinho. 2001. Loss of p16Ink4a with retention of p19Arf predisposes mice to tumorigenesis. Nature. 413:86-91.
- Shen, X., Y. Liu, Y.J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G.C. Yuan, and S.H. Orkin. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell. 32:491-502.
- Shi, Y., E. Seto, L.S. Chang, and T. Shenk. 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell. 67:377-88.
- Silva, J., W. Mak, I. Zvetkova, R. Appanah, T.B. Nesterova, Z. Webster, A.H. Peters, T. Jenuwein, A.P. Otte, and N. Brockdorff. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell. 4:481-95.

- Simon, J.A., and R.E. Kingston. 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol.
- Sing, A., D. Pannell, A. Karaiskakis, K. Sturgeon, M. Djabali, J. Ellis, H.D. Lipshitz, and S.P. Cordes. 2009. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell. 138:885-97.
- Singh, S.K., I.D. Clarke, M. Terasaki, V.E. Bonn, C. Hawkins, J. Squire, and P.B. Dirks. 2003. Identification of a cancer stem cell in human brain tumors. Cancer Res. 63:5821-8.
- Sleutels, F., R. Zwart, and D.P. Barlow. 2002. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 415:810-3.
- Sparmann, A., and M. van Lohuizen. 2006. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer. 6:846-56.
- Sridhar, V.V., A. Kapoor, K. Zhang, J. Zhu, T. Zhou, P.M. Hasegawa, R.A. Bressan, and J.K. Zhu. 2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. Nature. 447:735-8.
- Stewart, S.A., and R.A. Weinberg. 2006. Telomeres: cancer to human aging. Annu Rev Cell Dev Biol. 22:531-57.
- Stock, J.K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A.G. Fisher, and A. Pombo. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol. 9:1428-35.
- Su, I.H., A. Basavaraj, A.N. Krutchinsky, O. Hobert, A. Ullrich, B.T. Chait, and A. Tarakhovsky. 2003. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. Nat Immunol. 4:124-31.
- Su, I.H., M.W. Dobenecker, E. Dickinson, M. Oser, A. Basavaraj, R. Marqueron, A. Viale, D. Reinberg, C. Wulfing, and A. Tarakhovsky. 2005. Polycomb group protein ezh2 controls actin polymerization and cell signaling. Cell. 121:425-36.
- Sui, Y., Z. Yang, S. Xiong, L. Zhang, K.L. Blanchard, S.C. Peiper, W.S. Dynan, D. Tuan, and L. Ko. 2007. Gene amplification and associated loss of 5' regulatory sequences of CoAA in human cancers. Oncogene. 26:822-35.
- Suzuki, M., Y. Mizutani-Koseki, Y. Fujimura, H. Miyagishima, T. Kaneko, Y. Takada, T. Akasaka, H. Tanzawa, Y. Takihara, M. Nakano, H. Masumoto, M. Vidal, K. Isono, and H. Koseki. 2002. Involvement of the Polycomb-group gene

Ring1B in the specification of the anterior-posterior axis in mice. Development. 129:4171-83.

- Swigut, T., and J. Wysocka. 2007. H3K27 demethylases, at long last. Cell. 131:29-32.
- Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 126:663-76.
- Takihara, Y., D. Tomotsune, M. Shirai, Y. Katoh-Fukui, K. Nishii, M.A. Motaleb, M. Nomura, R. Tsuchiya, Y. Fujita, Y. Shibata, T. Higashinakagawa, and K. Shimada. 1997. Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. Development. 124:3673-82.
- Tanuma, N., S.E. Kim, M. Beullens, Y. Tsubaki, S. Mitsuhashi, M. Nomura, T. Kawamura, K. Isono, H. Koseki, M. Sato, M. Bollen, K. Kikuchi, and H. Shima. 2008. Nuclear inhibitor of protein phosphatase-1 (NIPP1) directs protein phosphatase-1 (PP1) to dephosphorylate the U2 small nuclear ribonucleoprotein particle (snRNP) component, spliceosome-associated protein 155 (Sap155). J Biol Chem. 283:35805-14.
- Terranova, R., S. Yokobayashi, M.B. Stadler, A.P. Otte, M. van Lohuizen, S.H. Orkin, and A.H. Peters. 2008. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell. 15:668-79.
- Tetsu, O., H. Ishihara, R. Kanno, M. Kamiyasu, H. Inoue, T. Tokuhisa, M. Taniguchi, and M. Kanno. 1998. mel-18 negatively regulates cell cycle progression upon B cell antigen receptor stimulation through a cascade leading to c-myc/cdc25. Immunity. 9:439-48.
- Tie, F., T. Furuyama, J. Prasad-Sinha, E. Jane, and P.J. Harte. 2001. The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development. 128:275-86.
- Tie, F., C.A. Stratton, R.L. Kurzhals, and P.J. Harte. 2007. The N terminus of Drosophila ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. Mol Cell Biol. 27:2014-26.

- Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato, J. Valcarcel, and R. Guigo. 2009. Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol. 16:996-1001.
- Tiwari, V.K., L. Cope, K.M. McGarvey, J.E. Ohm, and S.B. Baylin. 2008. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. Genome Res. 18:1171-9.
- Tokimasa, S., H. Ohta, A. Sawada, Y. Matsuda, J.Y. Kim, S. Nishiguchi, J. Hara, and Y. Takihara. 2001. Lack of the Polycomb-group gene rae28 causes maturation arrest at the early B-cell developmental stage. Exp Hematol. 29:93-103.
- Tolhuis, B., E. de Wit, I. Muijrers, H. Teunissen, W. Talhout, B. van Steensel, and M. van Lohuizen. 2006. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat Genet. 38:694-9.
- Tonkin, E., D.M. Hagan, W. Li, and T. Strachan. 2002. Identification and characterisation of novel mammalian homologues of Drosophila polyhomeoticpermits new insights into relationships between members of the polyhomeotic family. Hum Genet. 111:435-42.
- Trimarchi, J.M., B. Fairchild, J. Wen, and J.A. Lees. 2001. The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. Proc Natl Acad Sci U S A. 98:1519-24.
- Trinkle-Mulcahy, L., S. Boulon, Y.W. Lam, R. Urcia, F.M. Boisvert, F. Vandermoere, N.A. Morrice, S. Swift, U. Rothbauer, H. Leonhardt, and A. Lamond. 2008. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol. 183:223-39.
- Ulke-Lemee, A., L. Trinkle-Mulcahy, S. Chaulk, N.K. Bernstein, N. Morrice, M. Glover, A.I. Lamond, and G.B. Moorhead. 2007. The nuclear PP1 interacting protein ZAP3 (ZAP) is a putative nucleoside kinase that complexes with SAM68, CIA, NF110/45, and HNRNP-G. Biochim Biophys Acta. 1774:1339-50.
- Utama, B., D. Kennedy, K. Ru, and J.S. Mattick. 2002. Isolation and characterization of a new nucleolar protein, Nrap, that is conserved from yeast to humans. Genes Cells. 7:115-32.
- Vakoc, C.R., S.A. Mandat, B.A. Olenchock, and G.A. Blobel. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell. 19:381-91.

- Valk-Lingbeek, M.E., S.W. Bruggeman, and M. van Lohuizen. 2004. Stem cells and cancer; the polycomb connection. Cell. 118:409-18.
- van der Knaap, J.A., B.R. Kumar, Y.M. Moshkin, K. Langenberg, J. Krijgsveld, A.J. Heck, F. Karch, and C.P. Verrijzer. 2005. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Mol Cell. 17:695-707.
- van der Lugt, N.M., J. Domen, K. Linders, M. van Roon, E. Robanus-Maandag, H. te Riele, M. van der Valk, J. Deschamps, M. Sofroniew, M. van Lohuizen, and et al. 1994. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. Genes Dev. 8:757-69.
- van der Vlag, J., and A.P. Otte. 1999. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nat Genet. 23:474-8.
- van Lohuizen, M., M. Tijms, J.W. Voncken, A. Schumacher, T. Magnuson, and E. Wientjens. 1998. Interaction of mouse polycomb-group (Pc-G) proteins Enx1 and Enx2 with Eed: indication for separate Pc-G complexes. Mol Cell Biol. 18:3572-9.
- van Lohuizen, M., S. Verbeek, B. Scheijen, E. Wientjens, H. van der Gulden, and A. Berns. 1991. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. Cell. 65:737-52.
- Varambally, S., S.M. Dhanasekaran, M. Zhou, T.R. Barrette, C. Kumar-Sinha, M.G. Sanda, D. Ghosh, K.J. Pienta, R.G. Sewalt, A.P. Otte, M.A. Rubin, and A.M. Chinnaiyan. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature. 419:624-9.
- Vincenz, C., and T.K. Kerppola. 2008. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proc Natl Acad Sci U S A. 105:16572-7.
- Vire, E., C. Brenner, R. Deplus, L. Blanchon, M. Fraga, C. Didelot, L. Morey, A. Van Eynde, D. Bernard, J.M. Vanderwinden, M. Bollen, M. Esteller, L. Di Croce, Y. de Launoit, and F. Fuks. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 439:871-4.
- Voncken, J.W., H. Niessen, B. Neufeld, U. Rennefahrt, V. Dahlmans, N. Kubben, B. Holzer, S. Ludwig, and U.R. Rapp. 2005. MAPKAP kinase 3pK

phosphorylates and regulates chromatin association of the polycomb group protein Bmi1. J Biol Chem. 280:5178-87.

- Voncken, J.W., B.A. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps, and M. van Lohuizen. 2003. Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. Proc Natl Acad Sci U S A. 100:2468-73.
- Voncken, J.W., D. Schweizer, L. Aagaard, L. Sattler, M.F. Jantsch, and M. van Lohuizen. 1999. Chromatin-association of the Polycomb group protein BMI1 is cell cycle-regulated and correlates with its phosphorylation status. J Cell Sci. 112 (Pt 24):4627-39.
- Vonlanthen, S., J. Heighway, H.J. Altermatt, M. Gugger, A. Kappeler, M.M. Borner, M. van Lohuizen, and D.C. Betticher. 2001. The bmi-1 oncoprotein is differentially expressed in non-small cell lung cancer and correlates with INK4A-ARF locus expression. Br J Cancer. 84:1372-6.
- Vousden, K.H., and C. Prives. 2009. Blinded by the Light: The Growing Complexity of p53. Cell. 137:413-31.
- Wang, E.T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. Nature. 456:470-6.
- Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R.S. Jones, and Y. Zhang. 2004a. Role of histone H2A ubiquitination in Polycomb silencing. Nature. 431:873-8.
- Wang, L., J.L. Brown, R. Cao, Y. Zhang, J.A. Kassis, and R.S. Jones. 2004b. Hierarchical recruitment of polycomb group silencing complexes. Mol Cell. 14:637-46.
- Wang, Y.J., and H.W. Brock. 2003. Polyhomeotic stably associates with molecular chaperones Hsc4 and Droj2 in Drosophila Kc1 cells. Dev Biol. 262:350-60.
- Whitcomb, S.J., A. Basu, C.D. Allis, and E. Bernstein. 2007. Polycomb Group proteins: an evolutionary perspective. Trends Genet. 23:494-502.
- Wierstra, I. 2008. Sp1: emerging roles--beyond constitutive activation of TATA-less housekeeping genes. Biochem Biophys Res Commun. 372:1-13.
- Yamaki, M., K. Isono, Y. Takada, K. Abe, T. Akasaka, H. Tanzawa, and H. Koseki. 2002. The mouse Edr2 (Mph2) gene has two forms of mRNA encoding 90- and 36-kDa polypeptides. Gene. 288:103-10.

- Yannoni, Y.M., M. Gaestel, and L.L. Lin. 2004. P66(ShcA) interacts with MAPKAP kinase 2 and regulates its activity. FEBS Lett. 564:205-11.
- Yu, W., D. Gius, P. Onyango, K. Muldoon-Jacobs, J. Karp, A.P. Feinberg, and H. Cui. 2008. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 451:202-6.
- Zeggini, E., M.N. Weedon, C.M. Lindgren, T.M. Frayling, K.S. Elliott, H. Lango, N.J. Timpson, J.R. Perry, N.W. Rayner, R.M. Freathy, J.C. Barrett, B. Shields, A.P. Morris, S. Ellard, C.J. Groves, L.W. Harries, J.L. Marchini, K.R. Owen, B. Knight, L.R. Cardon, M. Walker, G.A. Hitman, A.D. Morris, A.S. Doney, M.I. McCarthy, and A.T. Hattersley. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 316:1336-41.
- Zhang, H., A. Christoforou, L. Aravind, S.W. Emmons, S. van den Heuvel, and D.A. Haber. 2004. The C. elegans Polycomb gene SOP-2 encodes an RNA binding protein. Mol Cell. 14:841-7.
- Zhang, T., Y. Sun, E. Tian, H. Deng, Y. Zhang, X. Luo, Q. Cai, H. Wang, J. Chai, and H. Zhang. 2006. RNA-binding proteins SOP-2 and SOR-1 form a novel PcG-like complex in C. elegans. Development. 133:1023-33.
- Zhao, J., B.K. Sun, J.A. Erwin, J.J. Song, and J.T. Lee. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 322:750-6.
- Zhu, J., D. Woods, M. McMahon, and J.M. Bishop. 1998. Senescence of human fibroblasts induced by oncogenic Raf. Genes Dev. 12:2997-3007.